was cleaved with BamHI and XhoI. Vector: a modified
pBluescript KS(+) after bulk excision from Lambda ZAP II."
BASE COUNT 191 a 163 c 191 g 145 t
ORIGIN

Query Match 5.2%; Score 262.4; DB 9; Length 660;
Best Local Similarity 70.6%; Pred. No. 1.4e-58;
Matches 450; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

```
Qy     186 ccaqacctcaaqaaaacactqttcaqatqaaqacaaaqatqqcaactqqqatqcttqqqq 245
           || ||  || || |   | ||  ||||| ||||||   || |||  ||||||||| |||||
Db     164 CCTGAGTTCCAGGACTGCACGCTCAGAGGAAGACCGTGAAGGCCTCTGGGATGCCTGGGG 223

Qy     246 cqactqqaqtqactqctccqqaacctqtqqqqqqqqaqcatcatattctctqccqaqatq 305
           |   ||||| || || ||| | ||||||| || || || || ||||| ||| |  | ||
Db     224 CCCTTGGAGCGAGTGTTCCAGAACCTGTGGTGGGGGTGCCTCCTATTCCCTGAGACGCTG 283

Qy     306 tttqactqqaaaqqaattqtqaaaqqqcaqaacattcqqtacaaqacatgcaqcaatcatqa 365
             |||   | | ||  ||||||||   || |||| || | |||||||| |||  ||
Db     284 CCTGAGCAGCAAGAGCTGTGAAGGAGAAATATTCGATATAGAACATGCAGTAATGTGGA 343

Qy     366 ctqcqctccaqatqcaqaaaatttcaqaqccaqcaqtqctcaqcctacaatgatqtcca 425
           |||||| ||||| |||  |||||| ||| || |||| || ||  | || || ||| |
Db     344 CTGGCCACCAGAAGCAGGTGATTTCCGAGCTCAACAGTGTTCTGCTCATAACGACGTCAA 403

Qy     426 qtatcaqqqqcattactatcaatqqcttccacqatataatqatcctqctqccccqtqtqc 485
           ||| || || ||    |||||||||||.||   || |||||| || |     || || |
Db     404 GTACCATGGACAGCTTTATGAATGGCTTCCTGTATCTAATGACCCGGACAATCCATGCTC 463

Qy     486 actcaaqtqtcatqcaaaaqqacaaaactttqqtqqtqqaqctqqcacctaaqqtactqqa 545
           |||||||||| || ||| |||||   | | |||| |||| || |||| |||||  | ||
Db     464 ACTCAAGTGCCAAGCAAAAGGAACCAGCCTGGTTGTGGAACTAGCACCGAAGGTCTTAGA 523

Qy     546 tqqaactcqttqcaacacqqactcctqqacatqtqtatcaqtqqcatctqtcaqqcaqt 605
           ||| || || ||| | || || ||  |||| |||| |||| ||| | || ||    ||
Db     524 TGGTACCCGAGCGTATACAGAATCACTGGATATGTGCATCAGCGGCCTATGCCAAATTGT 583

Qy     606 qqqctqcqtatcqqcaactqtqqaaqcaatqccaaqqaqqacaactqtquaqtctqtqcccq 665
           |||||  ||||  || |||||  |||| || ||| || || |||||||||| | ||   |||
Db     584 TGGCTGCGATCACCAGCTGGGGAGCACTGTCAAAGAAGATAACTGTGGAATGTGCAACGG 643

Qy     666 cqatqqctccacctqc 681
           |||||||| || |||
Db     644 AGATGGCTCAACGTGC 659
```

RESULT 14
BF543437/c
LOCUS BF543437 408 bp mRNA linear EST 11-DEC-2000
DEFINITION UI-R-Y0-abq-a-03-0-UI.r1 UI-R-Y0 Rattus norvegicus cDNA clone UI-R-Y0-abq-a-03-0-UI 5', mRNA sequence.
ACCESSION BF543437
VERSION BF543437.1 GI:11644444
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciuromorphi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 408)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9665
Email: msoares@blue.weeg.uiowa.edu

cDNA Library Preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com) This clone is also available through the I.M.A.G.E. Consortium at LLNL (info@image.llnl.gov). IMAGE ID= 1740944
Seq primer: M13 Forward.
FEATURES Location/Qualifiers
source 1..408
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-Y0-abq-a-03-0-UI"
/clone_lib="UI-R-Y0"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-Y0 library is a subtracted library derived from an individually-tagged normalized whole-eye (minus the lens) library. The driver for the subtraction consisted of a pool of all previous libraries (UI-R-A0, UI-R-A1, UI-R-E0, UI-R-E1, UI-R-C0, and UI-R-C1). The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dT track which allows identification of the library of origin of a clone within the mixture. The subtracted library (UI-R-Y0) was constructed as follows: PCR amplified cDNA inserts from previous library clones from which 3' ESTs had been derived were used as a driver in a hybridization with the normalized whole-eye library in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the UI-R-Y0 library. This procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996)"
BASE COUNT 91 a 113 c 105 g 99 t
ORIGIN

Query Match 5.1%; Score 259.6; DB 10; Length 408;
Best Local Similarity 85.5%; Pred. No. 6.5e-58;
Matches 289; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

```
Qy       1 atqqctt-ctqqaqaqccctqqtqqatqctqataqqqatqqtctttcatqcactctccc 60
           |||||||  |||||   |  |||||||  |||||| | | |||  |||| |    ||||
Db     448 ATGGCTTTCTGGAAAGGATCCTGGTGGCTGCTGACACGAATGATCTTTACAGTCTCTGGT 289

Qy      61 ctccccqcaqaccacaqcaqtqaqaaaatctcctqqaqccctatttcctttccqaqtttqcactt 120
           |||| |||| |||||||  |||||  |||||||||||||||  ||||||||  |||||  |||||
Db     288 CTCCTGCAGGCCACAGCAGAGAAGTCTCCTGGAGCCTACTTCCTTCCTGAGTTCGCACTC 229

Qy     121 tctcctqaqqqaaqtttctqqaaaqacacaaccaqqqqqqcaqttcctcacttatcqctat 180
           ||||||||||||| |||||  |||||  || |||||||  |||||||||| || ||||||
Db     228 TCTCCTCAGGGAAGCTTCTAGAAGATACCACAGGGAACAGTTCCTCACATACCGCTAT 169

Qy     181 qatqaccaqacctcaaqaaaacactqttcaqatqaaqacaaaqatqqcaactqqqatqct 240
           |||||  ||||||||||||||  |||  || ||||||||||||||||||||||||||||||
Db     168 GATGATCAGACCTCAAGAAATGCTCACTCCGATGAAGACAAAGATGGCAACTGGGATGCC 109

Qy     241 tqqqqcqactqqaqtqactqctccqqacctqtqqqqqqaqqaqcatcatattctctqcqq 300
           |||||  |||||||||||||||||||||||||||||||||||||||||| || ||||| ||||
Db     108 TGGGGAGACTGGAGTGACTGCTCCCGGACCTGTGGGGGAGGAGCCTCCTATTCTTTGCGA 49

Qy     301 agatqtttqactqqaaqqaattqtqaaqqqcaqaacat 338
           |||||| |||||||||||| ||||||| ||||||||
Db      48 AGATGTCTGACTGGAAGGAACTGTGAAGGACAGAACAT 11
```

RESULT 15
BI523618
LOCUS BI523618 581 bp mRNA linear EST 29-AUG-2001
DEFINITION 603051757F1 NIH_MGC_122 Homo sapiens cDNA clone IMAGE:5201199 5',

```
                Taqami,M., Taqawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
                Muramatsu,M. and Hayashizaki,Y.
  TITLE         RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
  JOURNAL       Unpublished (2001)
COMMENT         Contact: Yoshihide Hayashizaki
                Laboratory for Genome Exploration Research Group, RIKEN Genomic
                Sciences Center(GSC), Yokohama Institute,
                The Institute of Physical and Chemical Research (RIKEN)
                1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
                Tel: 81 45 503-9222
                Fax: 81-45-503-9216
                Email: genome-res@gsc.riken.go.jp
                URL:http://genome.gsc.riken.go.jp/
                Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
                ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
                 Normalization and subtraction of cap-trapper-selected cDNAs to
                prepare full-length cDNA libraries for rapid discovery of new
                genes. Genome Res. 10 (10), 1617-1630 (2000)
                 Waqi,K., Fujiwake,S., Inoue,K., Kodama,Y., Iqawa,M., Ohara,E.,
                Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,I., Matsuura
                ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
                Hayashizaki,Y.
                 RIKEN integrated sequence analysis (RISA) system--384-format
                sequencing pipeline with 384 multicapillary sequencer. Genome Res.
                10 (11), 1757-1771 (2000)
                 Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
                ,Y. and Hayashizaki,Y.
                 Computer-based methods for the mouse full-length cDNA
                encyclopedia: real-time sequence clustering for construction of a
                nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
                 Kondo,S., Shinagawa,A., Saito,F., Kiyosawa,H., Yamanaka,I., Aizawa
                ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
                Hayashizaki,Y.
                 Computational Analysis of Full-Length Mouse cDNAs Compared with
                Human Genome Sequences. Mamm. Genome 12, 673-677 (2001)
                 Please visit our web site (http://genome.gsc.riken.go.jp) for
                further details
                e mouse tissues
FEATURES                Location/Qualifiers
     source             1..622
                        /organism="Mus musculus"
                        /db_xref="taxon:10090"
                        /clone="A330077P03"
                        /clone_lib="RIKEN full-length enriched, adult male spinal
                        cord"
                        /sex="male"
                        /tissue_type="spinal cord"
                        /dev_stage="adult"
                        /lab_host="DH10B"
                        /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                        prepared and sequenced in Mouse Genome Encyclopedia
                        Project of Genome Exploration Research Group in Riken
                        Genomic Sciences Center and Genome Science Laboratory in
                        RIKEN, Division of Experimental Animal Research in Riken
                        contributed to prepare mouse tissues. 1st strand cDNA was
                        primed with a primer [5'
                        GAGAGAGAGAAAGATCCAAGAGCTCTTTTTTTTTTTTTTTVN-3']. cDNA was
                        prepared by using trehalose thermo-activated reverse
                        transcriptase and subsequently enriched for full length by
                        cap-trapper. cDNA went through one round of normalization
                        to Rot = 10.0 and subtraction to Rot = 45.0. Second
                        strand cDNA was prepared with the primer adapter of
                        sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAAT-3']. cDNA was
                        cleaved with XhoI and BamHI. Vector: a
                        modified pBluescript KS(+) after bulk excision from Lambda
                        FIX II."
BASE COUNT      151 a     157 c     172 g     142 t
ORIGIN
```

```
  Query Match          5.8%;  Score 293.6;  DB 9;  Length 622;
  Best Local Similarity   69.1%;  Pred. No. 6.5e-67;
  Matches  401;  Conservative    0;  Mismatches  179;  Indels    0;  Gaps    0;

Qy      186 ccagacctcaagaaacactcgttcagatgaagacaaagatggcaactgggatgcttgggg 245
            || ||  || || |   | || ||||| ||||||   || ||| |||||||||| |||||
Db       25 CCTGAGTTCCAGGACTGCACGCTCAGAGGAAGACCGTGAAGGCCTCTGGGATGCCTGGGG 84

Qy      246 cgactggagtgactgctccccggacctgtgggggaggagcatcatattcttctgcggaaatg 305
            |   ||||| || || ||| | |||||||| || || || || |||| ||| |   | ||
Db       85 CCCTTGGAGCGAGTGTTCCAGAACCTGTGGTGGGGGTGCCTCCTATTCCTGAGACGCTG 144

Qy      306 tttgactggaaggaattgtgaaggggagaacattcggtacaggacatgcagcaatcatgg 365
            |||   | | ||  |||||||||   || ||||| || |  |||||||| |||   ||
Db      145 CCTGAGCAGCAAGAGCTGTCAAGGGAGAAATATTCGATATAGAACATGCAGTAATGTGGA 204

Qy      366 ctgccctccagatgcagaaggatttcaggaggcaggtgatgctaacatacaatgatgtcca 425
            |||||| ||||| |||| |||| | |||| || ||||| || ||  | | || || ||| |
Db      205 CTGCCCACCAGAAGCAGGTGATTTCCGAGCTCAACAGTGTTCTGCTCATAACGACGTCAA 264

Qy      426 gtatcaggggcattactatgaatggctgccacgatataatgatcctgctgccccgttgtgc 485
            ||| || || ||    |||||||||||||||   || |||||| || |     || || |
Db      265 GTACCATGGACAGCTTTATGAATGGCTTCCTGTATCTAATGACCCCGACAATCCATGCTC 324

Qy      486 actcaagtgtcatgcacaaggacaaaacttggtggtggagctggcacctaaggtactgga 545
            |||||||| || ||| |||||   | | |||| |||| || ||||| ||||  | ||
Db      325 ACTCAAGTGCCAAGCAAAAGGAACCAGCCTGGTTGTGGAACTAGCACCGAAGGTCTTAGA 384

Qy      546 tggaactcgttgcaacacggactccttggacatgtgtatcagtggcatctgtcaggcagt 605
            ||| || || ||| | || || ||  |||| ||||| ||||| ||| | || ||    ||
Db      385 TGGTACGCGATGCTATACAGAATCACTGGATATGTGCATCAGCGGCCTATGCCAAATTGT 444

Qy      606 gggctgcgatcggcaactgggaagcaatgccaaggaggacaactgtggagtctgtgccgg 665
            ||||| |||||   || ||||| |||| || ||| ||||| ||||||||||| ||   |||
Db      445 TGGCTGTGATCACCAGCTGGGGAGCACTGTCAAAGAGGATAACTGTGGAGTGTGCCACGG 504

Qy      666 cgattggctccacctgcaggcttgtacggggacaatcaaagtcacacgtttctcctgaaaa 725
            |||||||| || ||| |||| || || || | || || | ||| |   || ||| |   ||
Db      505 AGATGGCTCAACGTGCCGGCTGGTCCGAGGGCAGTATAAATCTCAGCTCTCTGCAAGTAA 564

Qy      726 aagagaagaaaaatgtaattgctgttcctttqqqaagtcqa 765
            |   || || |  ||  | ||  |||||  || ||  ||
Db      565 ATCGGATGACACCGTGGTCGCCATTCCTTACGGCAGCAGA 604
```

```
RESULT   8
AQ743958
LOCUS       AQ743958                 890 bp    DNA     linear   GSS 16-JUL-20
DEFINITION  HS_5504_A1_D01_T7A RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=1080 Col=1 Row=G, DNA sequence.
ACCESSION   AQ743958
VERSION     AQ743958.1  GI:5521480
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 890)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering.bac.htm)
```

```
            or from Research Genetics (info@resgen.com). BAC end Web Server:
            http://www.htgsc.washington.edu
            Plate: 1080  row: G  column: 1
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 890.
FEATURES             Location/Qualifiers
     source          1..890
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate:1080 Col:1 Row:G"
                     /clone_lib="RPCI-11 Human Male BAC Library"
                     /sex="male"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     Male blood DNA was isolated from one randomly chosen donor
                     and partially digested with a combination of EcoRI and
                     EcoRI Methylase. Size selected DNA was cloned into the
                     pBACe3.6 vector at EcoRI sites"
BASE COUNT       228 a    195 c    218 g    259 t
ORIGIN


  Query Match             5.6%;  Score 281.8;  DB 12;  Length 890;
  Best Local Similarity   97.6%;  Pred. No. 1.1e-63;
  Matches  286;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy    1696 ccaacgttcattccagaaccctggtcagcctgcagtaccacgtgtgggccgggtgtggcag 1755
           || | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     125 CCTAGGTTCATTCCAGAACCCTGGTCAGCCTGCAGTACCACGTGTGGGCCGGGTGTGGCAG 184

Qy    1756 gtccgccgaggtgaagtgccgtgtgctcctcacattcaccgcagactgagactgagctgccc 1815
           ||||| |||||||||||||||||||||||||||||||||||||||||| | ||||||||||
Db     185 GTCCGTGAGGTGAAGTGCCGTGTGCTCCTCACATTCACCGCAGACTGATATTGAGCTGCCC 244

Qy    1816 gaggaagagtgtgaaggccccaagctgcccaccgaacggccctgcctcctggaaucatgt 1875
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     245 GAGGAAGAGTGTGAAGGCCCCAAGCTGCCCACCGAACGGCCCTGCCTCCTGGAAGCATGT 304

Qy    1876 gatgaagagccccggcctccccgauagctagacatccctctccctgaggacagtgaagacgact 1935
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     305 GATGAGAGCCCGGCCTCCCCGAGAGCTAGACATCCCTCTCCCTGAGGACAGTGAGACGACT 364

Qy    1936 tacgactgggagtacgctgggttcaccccttgcacagcaacatgcttgggagg 1988
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     365 TACGACTGGGAGTACGCTGGGTTCACCCCTTGCACAGCAACATGCTTGGGAGG 417


RESULT   9
BE888902
LOCUS       BE888902                 850 bp    mRNA    linear   EST 20-OCT-2000
DEFINITION  601513828F1 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:3915294 5',
            mRNA sequence.
ACCESSION   BE888902
VERSION     BE888902.1  GI:10445670
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 850)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov.
  TITLE     National Institutes of Health Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM0738  row: q  column: 07
            High quality sequence stop: 742.
FEATURES             Location/Qualifiers
     source          1..850
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3915294"
                     /clone_lib="NIH_MGC_71"
                     /tissue_type="leiomyosarcoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally   Primer: Oligo dT.
                     Average insert size 2.1 kb. "
BASE COUNT       202 a    219 c    228 g    201 t
ORIGIN


  Query Match             5.5%;  Score 278;  DB 10;  Length 850;
  Best Local Similarity   97.3%;  Pred. No. 1.1e-62;
  Matches  451;  Conservative    0;  Mismatches  215;  Indels    4;  Gaps    4;

Qy     186 ccagacctcaagaaacactcgttcagatgaagacaaaqatggcaactgggatgcttgggg 245
           || || || || | | | || || || || ||| || |||  ||||||||| ||||||
Db     103 CCTGAGTTCCAGGACCGCACGCTCCGAGGAGGACCGGGACGGCCTATGGGATGCCTGGGG 162

Qy     246 cgactgggatgactgctccgaacctgtgggggaggaacatcatattctctgcggaqatg 305
           |  ||||||| |||||| || ||||| ||||| || || || || || |||||| || | ||
Db     163 CCCATGGAGTGAATGCTCAGGCACCTGCGGGGGTGGGGCCTCCTACTCTCTGAGGCGCTG 222

Qy     306 tttgactgcaaqcaattgtgaaggccagaacatteggtacaaqacatgcagcaatcatga 365
           |||  | | ||  |||||||  || || || ||||  |||||||| |||  ||
Db     223 CCTGAGCAGCAAGACCTCTGAAGTAAGAAATATCCGATACAGAACATGCAGTAATGTGGA 282

Qy     366 ctgcccctccagatgccagaaqattttcgtgagccagcagtgctcaqcctacaatgatgtcca 425
           |||||| ||||| |||| ||||.| |||| |||| |||||||  | |||||||| |
Db     283 CTGCCCACCAGAAGCAGGTGATTTCCGACCTCAGCAATGCTCAGCTCATAATGATGTCAA 342

Qy     426 gtatcaqqqqcattactatgaatqqcttccgatataatgatcctgctgcccctgtgtgc 485
           | | || || | | |||||||||||||  | |||||| |||    ||| ||| |
Db     343 GCACCATGGGCAGTTTTATGAATGGCTTCCTGTGTCTAATGACCCTGACAACCCATGTTC 402

Qy     486 actcaaqtgtcatgcaaaqqaaaaaacttqgtqgtqqaqctqqcacctaaqqtactgga 545
           |||||||| || ||  ||||  || | |||| || || || ||||||||||  | ||
Db     403 ACTCAAGTGCCAAGCCAAAGGAACAACCCTGGTTGTTGAACTAGCACCTAAGGTCTTAGA 462

Qy     546 tqqaactqttqcaacacqqactcctqqacatqtqtatcaqtqqcatctqtcaqqcaqt 605
           ||| || || ||| | || | |||| |||| |||||||  | || ||    ||
Db     463 TGGTACGG-TGCTATACAGAATCTTTGGATATGTGCATCACTGGTTTATGCCAAATTGT 521

Qy     606 qqqctqcqatcqqcaactqqqaaqcaatqccaaqqaqqacaactqtqqaqtctqtqccqq 665
           ||||||||||  || ||||||||||  | |||||| || |||||| |||||   |||
Db     522 TGGCTGCGATCACCAGCTGGGAAGCACGTCAAGGAAGATAACTGTCGGGTCTGCAACGG 581

Qy     666 cqatqaqtccacctqcaqqcttqtacqqqqacaatcaaaqtcacacqtttc tcctqaa 723
           ||| || |||||||| |||| || || || || |  || || ||  | ||  |
Db     582 AGATGGGGTCCACCTGCCGGCTGGTCCGAGGGCAGTATAAATCCCAGCTCTCAGCAACCC 641

Qy     724 aaaaqaqaaqaaaatqtaattqc-tqttcctttqqaaqtcqaaqtqtqaqaattacaqt 782
           ||  | || || |||  ||||   |||| |  || |     | | |     ||  |
Db     642 AATCGGGATGATACTGTGGTTGCAAATTCCCTATGGGAAGTTGACTATTTCGCTTGGTTT 701

Qy     783 ttaaaaactq  c t  t  ttttattgaatcaaaaaqacttcaaaaaaaaaaaaqaaqca 842
           ||||| |||  ||| | ||  | | | | |||||       ||| ||| |  | |
Db     702 AAAAGGGCTGATCACTTTTTATTGGGAACCAAAACCTGCACGGGACTCAAGGTGGCAAG 761

Qy     843 cagctttaac 852
             | | || |
Db     762 AGGGTCTATC 771


RESULT  10
BB478282
LOCUS       BB478282                 619 bp    mRNA    linear   EST 25-OCT-2001
DEFINITION  BB478282 RIKEN full-length enriched, 13 days embryo heart Mus
```

GenCore version 4.5
Copyright (c) 1993 - 2001 Compugen Ltd.

SW protein - protein search, using sw model

Run on: August 19, 2002, 5:44:11 ; Search time 5.12 Seconds
(without alignments)
4166.415 Million cell updates/sec

Title: US-09-784-358-2
Perfect score: 9446
Sequence: 1 MASWTSPWWVLLGMVFMHSP.....LNLGSLDRYKQKQQQSPQES 1001

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747544 seqs, 113027706 residues

Total number of hits satisfying chosen parameters: 747544

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: A Geneseq 032802:*
1: /cgn2_6/ptodata/2/geneseq/geneseq_pep_2:*
2: /cgn2_6/ptodata/2/geneseq/geneseq_pep_3:*
3: /cgn2_6/ptodata/2/geneseq/geneseq_pep_4:*
4: /cgn2_6/ptodata/2/geneseq/geneseq_pep_5:*
5: /cgn2_6/ptodata/2/geneseq/geneseq_pep_6:*
6: /cgn2_6/ptodata/2/geneseq/geneseq_pep_7:*
7: /cgn2_6/ptodata/2/geneseq/geneseq_pep_8:*
8: /cgn2_6/ptodata/2/geneseq/geneseq_pep_9:*
9: /cgn2_6/ptodata/2/geneseq/geneseq_pep_10:*
10: /cgn2_6/ptodata/2/geneseq/geneseq_pep_11:*
11: /cgn2_6/ptodata/2/geneseq/geneseq_pep_12:*
12: /cgn2_6/ptodata/2/geneseq/geneseq_pep_13:*
13: /cgn2_6/ptodata/2/geneseq/geneseq_pep_14:*
14: /cgn2_6/ptodata/2/geneseq/geneseq_pep_15:*
15: /cgn2_6/ptodata/2/geneseq/geneseq_pep_16:*
16: /cgn2_6/ptodata/2/geneseq/geneseq_pep_17:*
17: /cgn2_6/ptodata/2/geneseq/geneseq_pep_18:*
18: /cgn2_6/ptodata/2/geneseq/geneseq_pep_19:*
19: /cgn2_6/ptodata/2/geneseq/geneseq_pep_20:*
20: /cgn2_6/ptodata/2/geneseq/geneseq_pep_21:*
21: /cgn2_6/ptodata/2/geneseq/geneseq_pep_22:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	9446	100.0	1691	22	AAE07863	Novel human protei
2	9439	99.9	2691	22	AAE10859	Human novel KIAA12
3	8882	94.0	1627	22	AAE07870	Novel human protei
4	7066	74.8	1281	22	AAE10610	Human novel KIAA12
5	7060	74.7	1281	22	AAE10608	Human novel KIAA12
6	5086	53.8	926	22	AAB83370	NOVX4 protein seqpt
7	4604	48.7	886	21	AAB40500	Human ORFX ORF264
8	4480	47.4	845	22	AAE07808	Novel human protei
9	4225	44.7	771	22	AAE07869	Novel human protei
10	4056	42.9	724	22	AAE07866	Novel human protei
11	3861	40.9	1235	22	AAB06541	Human secreted pro
12	3860.5	40.9	1760	22	AAB82748	Human cDNA SEQ ID
13	3442	36.4	650	22	AAE07867	Novel human protei
14	2891	30.6	874	21	AAB41431	Human IAN0 234 (f
15	2841.5	30.1	615	21	AAY87347	Human signal pepti
16	2344	24.8	444	22	AAE07864	Novel human protei
17	2258	23.9	525	2	AAY41720	Human PRO1071 prot
18	2258	23.9	525	22	AAB29241	Human PRO polypept
19	2251	23.8	525	21	AAB44285	Human PRO1071 (UNQ
20	2048	21.6	525	22	AAB72200	Human ADAM15 RI am
21	1880	20.1	572	22	AAE07865	Novel human protei
22	1560	16.7	281	22	AAB01315	Human BSPP 124 pro
23	1536.5	16.4	428	21	AAB41561	Human ORFX ORF162
24	1486	15.9	417	22	AAG78679	Human thrombotic p
25	1362	14.6	480	21	AAB01425	Human IAN00 224.
26	1183.5	12.7	2150	21	AAY53898	Amino acid sequenc
27	1183.5	12.7	2169	22	AAB90017	Human secreted pro
28	1043	11.2	1934	22	AAB72201	Human ADAM15 C alt
29	1028	11.0	1882	22	AAB72206	Human ADAM15 C ami
30	938.5	10.0	262	22	AAB90582	Human secreted pro
31	917.5	9.8	1784	21	AAB41376	Human ORFX ORF143
32	901	9.6	1600	22	AAB80644	Human metalloprote
33	897	9.6	1686	21	AAB41944	Human ADAM type se
34	892	9.5	1686	22	AAE30913	Human 27825 ADAM 1
35	891	9.5	1686	22	AAE30944	Human 27825 ADAM 1
36	891	9.5	2858	22	AAB71156	Drosophila melanog
37	891	9.5	3060	22	AAB58064	Drosophila melanog
38	828.5	8.8	1104	22	AAB74945	Human ADAM type me
39	816.5	8.6	1472	22	AAB72200	Human ADAM15 C sol
40	804	8.5	634	22	AAB01201	Human thrombospond
41	798.5	8.5	1081	22	AAB72208	Human ADAM15 lo am
42	774.5	8.2	858	22	AAB58082	Drosophila melanog
43	769	8.2	645	22	AAB72453	Human cDNA SEQ ID
44	767	8.1	603	21	AAB41128	Human PLIT/Eosin.
45	765	8.1	603	21	AAB41254	Human metalloprote

ALIGNMENTS

RESULT 1
AAE07863
ID AAE07863 standard; protein; 1691 AA.
XX
AC AAE07863;
XX
DT 01-NOV-2001 (first entry)
XX
DE Novel human protein (NHP) #1.
XX
KW Novel human protein; NHP; thrombospondin gene therapy; cancer;
KW cytostatic; antisense therapy; angiogenesis; hormonal disorder.
XX
OS Homo sapiens.
XX
PN WO200161011-A2.
XX
PD 23-AUG-2001.
XX
PF 15-FEB-2001; 2001WO-US05290.
XX
PR 17-FEB-2000; 2000US-0183282.
XX
PA (LEXI-) LEXICON GENETICS INC.
XX
PI Donoho G, Stecenille L, Turner CA, Friedrich G, Zambrowicz B;
PI Sands AT;
XX
DR WPI; 2001-514776/56.
DR N-PSDB; AAE14364.
XX
PT An isolated nucleic acid encoding a thrombospondin useful as a
PT hybridization probe and gene therapy treatments of cancer
XX

PS Claim 2; Page 31-34; 56pp; English
XX
CC The invention relates to novel human polynucleotides encoding proteins
CC that share sequence similarity with animal proteins having thrombospondin
CC repeats. NHPs are expressed in human cell lines, pituitary, lymph node,
CC prostrate, testis, adrenal gland, uterus, foetal kidney, foetal lung and
CC gene trapped human cells. NHPs are thrombospondins useful for treating
CC biological disorders involving angiogenesis, cancer and development and
CC also in pharmacogenomic applications. NHPs are useful as a hybridisation
CC probe for screening libraries, assessing gene expression patterns and
CC also in gene therapy. Proteins having thrombospondin repeats act as
CC receptors, secreted extracellular matrix proteins and proteases. The
CC present sequence is novel human protein (NHP).
XX
SQ Sequence 1691 AA;

Query Match 100.0%; Score 9346; DB 22; Length 1691;
Best Local Similarity 100.0%; Pred No. 0;
Matches 1691; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 MASWISPWWVLIGMVFMHSPIPQITAEKSPGAYFLPEFALSPQGSFLEDFTGEQPLIYRY 60
Db       1 maswtspwwvligmvfmhspipqittaekspgayflpefalspqgsfledftgeqplityry 60

Qy      61 DDQTSFNTFSTFMKTCNWFAWGPWSPCSRTCGGGASYSLRPCLTGRNCLGQNIEYKTCSN 120
Db      61 ddqtsrntfstfmktcnwfawgpwspcsrtcgggasysirpclsgrncigqnieykscsn 120

Qy     121 HDCPEPKAIEHRAQQTFAYTWVPKTPHYYIWSFFNDVAAFCAIFCMAFSCLNIVVEIAIFV 180
Db     121 hdcppdaedtraqqcsayndvqyqqhyycwlpryndpraapcalkchqrqgnlmvcialpkv 180

Qy     181 IPSIRTNIISIPMCIKCPYAVVSFMQLGSNAKELNKGVCAQESCPLVPCQSRSHVSP 240
Db     181 idqtrentdsldmcisqicqavscdrqlgsnakednegvcaqdqsterlvrgqgskshvsp 240

Qy     241 EKPPPNVIAVPLCSPSVPITVKCPAHLPTIQKTLQCSKCEHSPNSPCVPVVENTTVEPQR 300
Db     241 ekrcsn-tvqpltsr-rttkkgpahltlcskttlpgskpchafnsggvtrventtveiqr 300

Qy     301 GSEPQTIPKIPGPIMADPIPKTPYTAAKDSVVQFFFVGPISHQWPQTIPPPCTVTCCGCYQ 360
Db     301 gserqtfpipgpimadfiftrytaakdsvvqfffvgpishqwpqtipppctvtccgcyq 360

Qy     361 INSAFCVDTRLKRVVPDHYCHYYPENVKPKPKLKECSMDPCPSSDGPKEIMPYDHFQPLP 420
Db     361 lnsaccvdtrlkrvvpdhychyypenvkpkpklkecsmdpcpssdgpkeimpydhfqplp 420

Qy     421 RWIIHNEWTACDVSCGGGIQKRSFVCVPIPMHGFIIQVPEWKCMTAFEPEVMQTCNLFDCP 480
Db     421 rwchapwtacsvscggqiqkrsfvcvpipmhgfiiqvpewkcmtafepevmqtcnlfdcp 480

Qy     481 EWIAMEWSQCTVTCGKGLRYRVVLCINSEGEHVGCLNPSLRLPEECVIPECYRTKEK 540
Db     481 kwiamewsqctvtcgkglryrvvlcinsegehvgclnpslrlpeecvipecyrtkek 540

Qy     541 SPVPAKLPWIKQAQFLEFTRIATEEPTPIPEPWSACSITCGPGVQVREVKCRVLLIFIQT 600
Db     541 spveuklpwlkqaqflefirlateeptpipepwsacsitcgpgvqvrevkcrvllifiqt 600

Qy     601 ETELPEEECEGPKLPTERPCLLEACDESPASRELDIPLPEDSETTYDWEYAGPIPCTATC 660
Db     601 etelpeeeceqpklpterpclleacdespasreldiplpedsettydweyaqptpctatc 660

Qy     661 LGGHQPAIAVFMKDGQLTVNEVLRPMVHEPTAMSLAFNIFCLFFWHVSWGFIDATCC 720
Db     661 lqqhqpaiavfmkdgqltvnevlrpmvhepttamslafnifclffwhvswgfidatcc 720

Qy     721 VGIQIRDVYCLHPGETPAPPEECKDEKPHALQACNQFDCPPGWHIEEWQQCSRTCGGGTQ 780
Db     721 vgiqirdvyclhpgetpappeeckdekphalqacnqfdcppgwhieewqqcsrtcgggtq 780

Qy     781 NRPVECPQLITPCSFLNFSVELQQGPFASFHKSCAKIGCPPHLAVGDWSKCSVSCGVGIQ 840
Db     781 srrvtcrqlitpqsflnlsdelqqgpfasfhksckakigcpphlavgdwskcsvscgvgiq 840

Qy     841 FRKCVEQFLAAFCFFIFLSEMMCFDLFDFFLVFSQQMFETSKIKSLMETKLCEQFVILS 900
Db     841 rrkqrcqrlaakgrrifisemmcfdlfdffvfsqqmfetskiksimktklqeqfqils 900

Qy     901 VQSVVTQTPFEKPINLTDGSFAYLFPNTSVTIPCPVPFFQKSLQWEFTGSCIQNSERLS 960
Db     901 vqsvvtqtpfekpinltdgsfayllpntsvtipcpvpffqksiqweftgscidnserls 960

Qy     961 ITKSGSLKIHGLAAPNICVYPCIAGSAQLTVVLKLIGTDNELIAPPALFIPDMPEYPCMDH 1020
Db     961 itksgslkihglaapdicvypciagsaqltvvlkligtdneliappalfipdmpeypcmdh 1020

Qy    1021 SFANSLGVTWHKMRQMWNNKNDLYLDDDHISNQPFLPALLCHCSNGACSTNSWELKNKQF 1080
Db    1021 scanslgvtwhkmrqmwnnkndlyldddhisnqpflpallghcsnsaqstnswelknkqf 1080

Qy    1081 EAAVKQGAYSMDTAQFDFLIPNMSQLMETCFVSDDLASQLIYQLVAELAKAQPTHMQWRG 1140
Db    1081 eaavkqgaysmdtaqfdflipnmsqlmetcfvsddlasqliyqlvaelakaqpthmqwrq 1140

Qy    1141 TQFPTPPAAQLPGPTGSVSQSSHAKNSGKLTPKPKGPVLMRQSQPPSISTNKTINSRIGN 1200
Db    1141 iqeetppaaqlrgetgsvsqsshakusqkltikpkgpvlmrqsqppsistnktinsriqn 1200

Qy    1201 TVYITKPTFVINTLGDLITPSFATYTWTKDGTLLQPSVKTLLVGTGKTQTQNPTPPEQGT 1260
Db    1201 tvyitkrtevinllgdlitpscatytwtkdgtllqpsvkllldgtgkiqiqnptrkeqqi 1260

Qy    1261 YECSVANHLGSDVESSSVLYAEAPVILSVERNITKPPHNHLSVVVGGIVEAALGANVTIR 1320
Db    1261 yecsvanhlgsdvesssvlyaeapvilsvernitkpphnhlsvvvggivealganvtir 1320

Qy    1321 CPVKGVPQPNITWLKNGGSLSGNVSLLPNGSLLIQNVSLENEGTYVCIATNALGKAVATS 1380
Db    1321 cpvkgvpqpnitwlknggslsgnvsllpngsllignvslenegtyvciatnalgkavats 1380

Qy    1381 VLHLLERRWPESRIVFLQGHKKYILQATNTRTNSNDPTGEPPPQEPTPWEPGNWSHCSATC 1440
Db    1381 vlhllerrwpesrivflqghkkyilqatntrtnsndptgepppqeptpwepgnwshcsatc 1440

Qy    1441 GHLGARTQRPQCVMANGQEVSPEALCDHIQKPLAGEPPCNIRDCPARWFTSVWSQCSVSCG 1500
Db    1441 ghlgartqrpqcvmangqevspealcdhiqkplagepcnirdcparwftsvwsqcsvscg 1500

Qy    1501 EGYHSRQVECKRPDKANGTVQVVSPRACAPKDRPLQRKPCFGHPCVQWEPGNRCPGRCMGR 1560
Db    1501 egyhsrqveckrpdkangtvqvvspracapkdrplqrkpcfghpcvqwepgnrcpgrcmgr 1560

Qy    1561 AVRMQGEHTACQHNSSTCNPIDFRKRFTLFSNTLGACVCWDTSIWFPCTAACGPGPQSR 1620
Db    1561 avrmqgrhtacqhnssdcnpiddfrkrftlfsntlgacvcwdtsiwfpctaacgpgpqsr 1620

Qy    1621 KVPCTHTFSCKPVAKRHCVQKPKPISWPHCLGPSCPETCTPTHYCMPVKPLNLCSLGRY 1680
Db    1621 kvpcthtfsckpvakrhcvqkpkpiswphclgpscpetctpthycmpvkplnlcslgry 1680

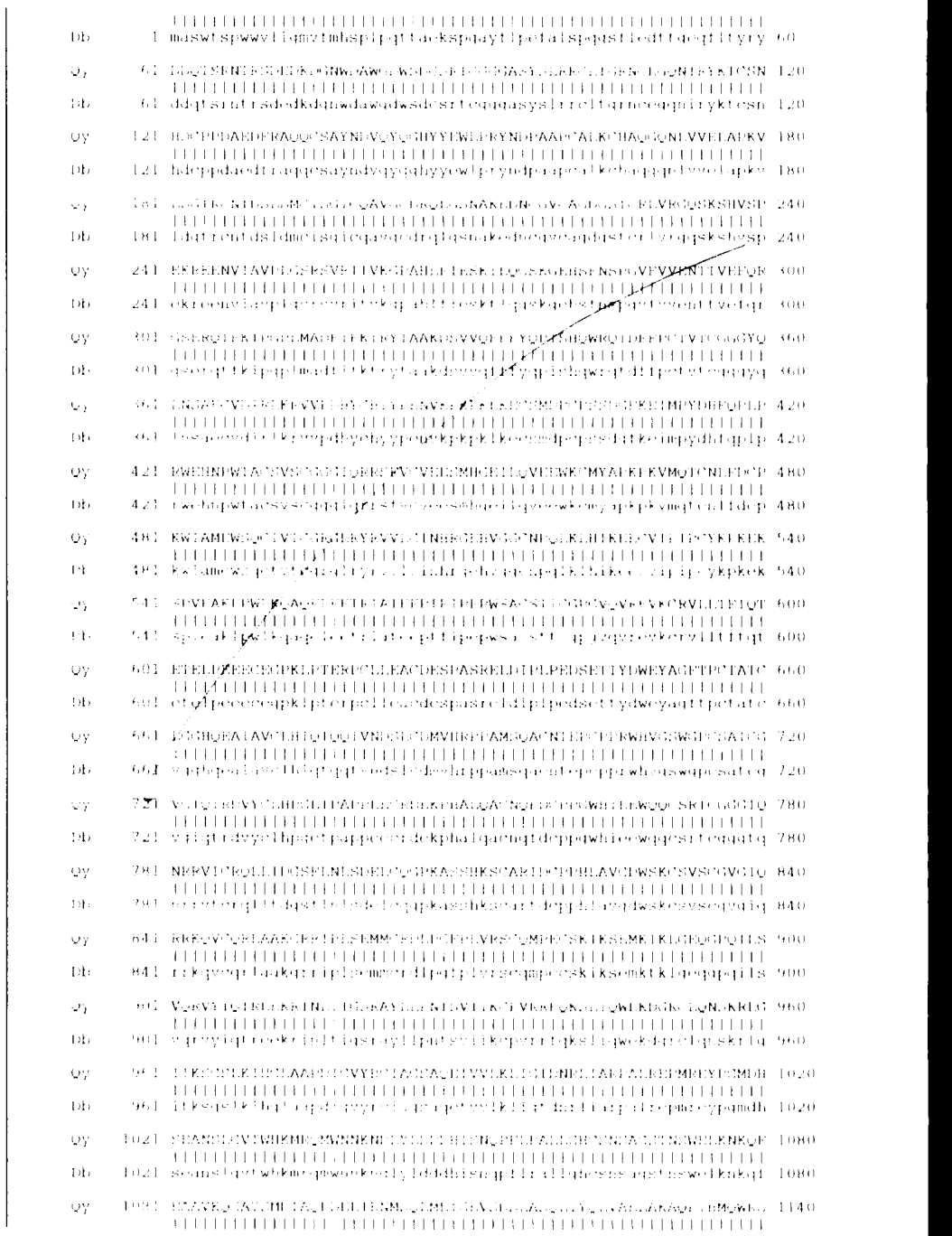
Qy    1681 KQRCCQSCQEG 1691
Db    1681 kqrccqscqeq 1691
```

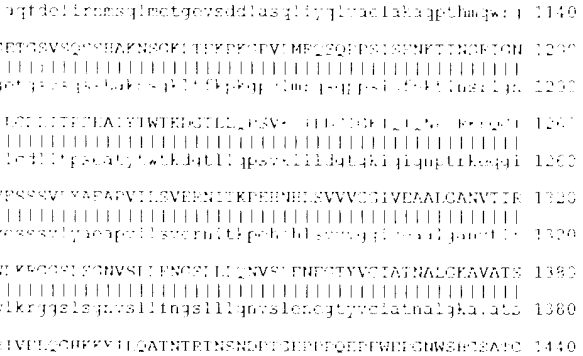
SELECT 2
AAE10609
ID AAE10609 standard; Protein; 1691 AA.
XX
AC AAE10609;
XX
DT 19-DEC-2001 (first entry)
XX

```
DE   Human novel KIAA1233 like protein, NOV 2a.
XX
KW   Human; NOV X protein; KIAA1233 like protein; STE20 like protein; tumour;
KW   trypsin inhibitor like protein; gene therapy; haematopoietic; cancer;
KW   immunological disorder; neurodegenerative disorder; Alzheimer's disease;
KW   Parkinson's disease; immunomodulatory; pharmacogenomic; haemostatic;
KW   human immunodeficiency virus; HIV; fertility disorder; neuroprotective;
KW   cytostatic; nootropic; anti infertility; cancer; chromosome 15;
KW   NOV 2a protein.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Active site     1..20
FT                   /label= Signal peptide
FT   Protein         27..1691
FT                   /note= "Human mature novel KIAA1233 like protein,
FT                   NOV 2a"
XX
PN   WO200162928-A2.
XX
PD   30-AUG-2001.
XX
PF   26-FEB-2001; 2001WO-US006751.
XX
PR   25-FEB-2000; 2000US-0184761.
PR   28-FEB-2000; 2000US-0185418.
PR   01-MAR-2000; 2000US-0185967.
PR   18-APR-2000; 2000US-0197724.
PR   27-APR-2000; 2000US-0199957.
PR   23-FEB-2001; 2001US-0789390.
XX
PA   (CURA ) CURAGEN CORP.
XX
PI   Vernet CAM, Fernandes E, Shimkets RA, Mandekupalil J, Spaderna SK;
XX
DR   WPI; 2001-582051/65.
DR   N-PSDB; AAD17757.
XX
PT   New isolated KIAA1233 like, STE20 like, or trypsin inhibitor like
PT   polypeptide for diagnosing and treating pathological disorders, such as
PT   Parkinson's disease and for use in pharmacogenomics
XX
PS   Claim 1; Page 22-29; 180pp; English.
XX
CC   The invention relates to novel human polypeptides referred as NOV X
CC   and their corresponding nucleic acid sequences. NOV X collectively
CC   include NOV 1, NOV 2a and NOV 2b which are novel KIAA1233 like
CC   polypeptides, NOV 3a, NOV 3b, NOV 3c and NOV 3d which are novel
CC   STE20 like polypeptides and NOV 4a, NOV 4b, NOV 4c, NOV 4d and NOV 4e
CC   which are novel trypsin inhibitor like polypeptides. NOV X is used to
CC   identify a potential therapeutic agent that can modulate its activity
CC   and can be used for treating a pathology related to aberrant expression
CC   or aberrant physiological interactions of NOV X. NOV X or its DNA is
CC   used to determine the presence or predisposition to a disease associated
CC   with altered levels of NOV X, NOV X, its DNA and its antibody, are used
CC   to treat or prevent a pathology associated with NOV X. The pathological
CC   states that can be treated or prevented are haematopoietic, cancer,
CC   immunological, tumour, neurodegenerative (e.g. Alzheimer's and
CC   Parkinson's diseases), bone, hemostatic (e.g. virus (HIV), illness and
CC   fertility disorders. NOV X and its DNA are used in pharmacogenomics for
CC   predictive medicine. NOV X DNA is used in gene therapy. The present
CC   sequence is human novel KIAA1233 like protein, NOV 2a. NOV 2a gene is
CC   located at chromosome 15.
XX
SQ   Sequence  1691 AA;
```

Query Match 99.7%; Score 4439; DB 22; Length 1691;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1688; Conservative 1; Mismatches 1; Indels 1; Gaps 1;

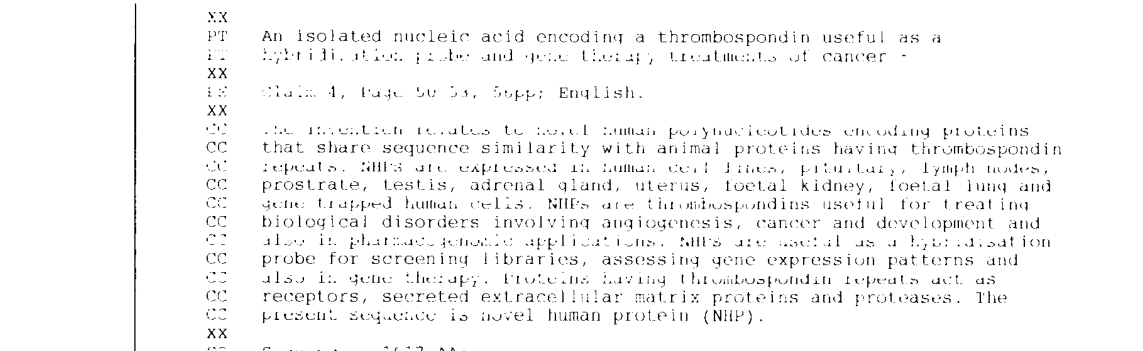
```
Qy       1 MASWTSPWWVLQMVTMHSPLPQTTAEKSPQAYELPEFALSPQQSFLEDCTGRQFTYRY 60
```





RESULT 3
AAE07870
ID AAE07870 standard; Protein; 1617 AA.
XX
AC AAE07870;
XX
DT 01-NOV-2001 (first entry)
XX
DE Novel human protein (NHP) #8
XX
KW Novel human protein; NHP; thrombospondin; gene therapy, cancer,
KW cytostatic; antisense therapy; angiogenesis; biological disorder.
XX
OS Homo sapiens.
XX
PN WO200161011-A2.
XX
PD 23-AUG-2001.
XX
PF 15-FEB-2001; 2001WO-US05290.
XX
PR 17-FEB-2000; 2000US-0183282.
XX
PA (LEXI-) LEXICON GENETICS INC.
XX
PI Donoho G, Scoville J, Turner CA, Friedrich G, Zambrowicz B;
PI Sands AT;
XX
DR WPI; 2001-514776/56.
DR N-PSDB; AAD14471.

XX
PT An isolated nucleic acid encoding a thrombospondin useful as a
PT hybridisation probe and gene therapy, treatments of cancer -
XX
PS Claim 4; Page 50-53; 56pp; English.
XX
CC The invention relates to novel human polynucleotides encoding proteins
CC that share sequence similarity with animal proteins having thrombospondin
CC repeats. NHPs are expressed in human cell lines, pituitary, lymph nodes,
CC prostrate, testis, adrenal gland, uterus, foetal kidney, foetal lung and
CC gene trapped human cells. NHPs are thrombospondins useful for treating
CC biological disorders involving angiogenesis, cancer and development and
CC also in pharmacogenomic applications. NHPs are useful as a hybridisation
CC probe for screening libraries, assessing gene expression patterns and
CC also in gene therapy. Proteins having thrombospondin repeats act as
CC receptors, secreted extracellular matrix proteins and proteases. The
CC present sequence is novel human protein (NHP).
XX
SQ Sequence 1617 AA;





RESULT 4
AAE10610
ID AAE10610 standard; Protein; 1281 AA.
XX
AC AAE10610;
XX
DT 10-DEC-2001 (first entry)
XX
DE Human novel KIAA1233 like protein, NOV-2b.
XX
KW Human; NOV X protein; KIAA1233 like protein; C1B22 like protein; tumour;
KW trypsin inhibitor like protein; gene therapy; haematopoietic; illness;
KW immunological disorder; neurodegenerative disorder; Alzheimer's disease;
KW Parkinson's disease; immunomodulatory; pharmacogenomic; haemostatic;
KW human immunodeficiency virus; HIV; fertility disorder; contraceptive;
KW cytostatic; nootropic; anti infertility; cancer; chromosome 15;
KW NOV 2b protein.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Active site 1..26
FT /label= Signal_peptide
FT Protein 27..1281
FT /note= "Human mature novel KIAA1233-like protein,
FT NOV-2b"
XX
PN WO200162928-A2.
XX
PD 30-AUG-2001.
XX
PF 26-FEB-2001; 2001WO-US06151.
XX
PR 25-FEB-2000; 2000US-0184951.
PR 28-FEB-2000; 2000US-0185548.
PR 01-MAR-2000; 2000US-0185967.
PR 28-APR-2000; 2000US-0197723.
PR 27-APR-2000; 2000US-0199957.
PR 23-FEB-2001; 2001US-0789590.
XX
PA (CURA) CURAGEN CORP.
XX
PI Vernet CAM, Fernandes E, Shimkets RA, Macdougall J, Spaderna DK;
XX
DR WPI; 2001-5[illegible]-[illegible].
DR N-PSDB; AAD17758.
XX
PT New isolated KIAA1233-like, STE20-like, or trypsin inhibitor like
PT polypeptide for diagnosing and treating pathological disorders, such as
PT Parkinson's disease and for use in pharmacogenomics
XX
PS Claim 1; Page 37; 189pp; English.
XX
CC The invention relates to novel human polypeptides referred as NOV X
CC and 2b (i.e. [illegible]) as well as [illegible] NOV X collectively
CC include NOV 1, NOV 2a and NOV 2b which are novel KIAA1233-like
CC polypeptides, NOV 3a, NOV 3b, NOV 3c and NOV 3d which are novel
CC STE20-like polypeptides and NOV 4a, NOV 4b, NOV 4c, NOV 4d and NOV 4e
CC which are novel trypsin inhibitor like polypeptides. NOV X is used to
CC identify a potential therapeutic agent that can modulate its activity
CC and can be used for treating a pathology related to aberrant expression
CC or aberrant physiological interactions of NOV X. NOV X or its DNA is
CC used to determine the presence or predisposition to a disease associated
CC with altered levels of NOV X. NOV X, its DNA and its antibody are used
CC to treat or prevent a pathology associated with NOV X. The pathological
CC states that can be treated or prevented are haematopoietic, cancer,
CC immunological, tumour, neurodegenerative (e.g. Alzheimer's and
CC Parkinson's disease), human immunodeficiency virus (HIV) illness and
CC fertility disorders. NOV X and its DNA are used in pharmacogenomics for
CC predictive medicine. NOV X DNA is used in gene therapy. The present
CC sequence is human novel KIAA1233 like protein, NOV 2b. NOV 2b gene is
CC located at chromosome 15.
XX
SQ Sequence 1281 AA;

Query Match 75.6%; Score 7066; DB 22; Length 1281;

Best Local Similarity 94.7%, Pred. No. 0,
Matches 1280; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 411 MPYDHPQPLPRWEHNPWTACSVSPQGGIQKRSFVCVEESMHGEILQVEEWKCMYAPKPKV 470
Db 1 mpydhtrgplprwehnpwtarsvspqggiqkrsfvcveesmhgeilqveewkcmyapkpkv 60

Qy 471 MQTCNLFIMLPRWIAMEWSQCIVICLHGLRYFVVLICNHRGFHVGGNPQLIKHIKEECVI 530
Db 61 mqtcnlfdmlprwiamewspcfvicrhglryfvvlicnhrgfhvggnpqlikhikeecvi 120

Qy 531 PIPCYKPKKSSPVEAKLPWLPQAQELERLRIATEEPIFIPEPWSAQSTIQGPGVQVPPVK 590
Db 121 plpcykpkksspveaklpwlpqaqelerlriateepifipepwsaqstiqgpgvqvppvk 180

Qy 591 CRVLLTFTQTPTELPEEECEGPKLPTERPCLLEACDESPASREIDIPLPEDSETTYDWEY 650
Db 181 crvlltftqtpteipeeeceqpklptterpcilleacdespasreidiplpedsettydwey 240

Qy 651 AGPIPCIAICLGGHQEAIAVCLHIQTGQIVNDSLCDMVHRPAMSQACNTPPCPPRWHVG 710
Db 241 agpipciaiclgghqeaiavclhiqtgqtvndslcdmvhrppamsqacntppcpprwhvg 300

Qy 711 SWGPCSATCGVGIQTRDVYCLHPGETDAPPEECRDEKPHAIQACNQFDCPPGWHIEEWQQ 770
Db 301 swgpcsatcgvgiqtrdvyclhpgetdappeecrdekphaiqacnqfdcppgwhieewqq 360

Qy 771 CSEPCVGGTQNRSVICRLLDGSTENSCSELPKPEHAVPCHLPECAVTEWSP 821
Db 361 csrpcqggtqnrsvtcrqlddstnrsdcqqpkrshrarpgpvrdtaes 411

Qy 822 CSVSCGVGIQFRSVQCRLAAGCSKTTLSQSMCRGLGCPIVKCQLMGQKKFSLEMPIK 880
Db 412 csvscgvgiqfrsvqcrlaagcskttlsqsmcrglgcpivkcqlmgqkkfslempik 470

Qy 881 KDKQKCIPPKPKIKPLGPVSGMKLMILMLLPSVQGDSMETSDQKSVSEDQECLAVCIE 940
Db 471 kdkqkcippkpkikplgpvsgmklmilmllpsvqgdsmetsdqksvsedqeclavcie 530

Qy 941 RCILQNPKKLTTTKSCSKLKICHTAAPQGVYKCIACSAQIHVVLKLTGTNKLIAEFAIPE 1000
Db 531 rcilqnpkklttskscskliichtaapqgvykciacsaqihvvlkltgtnklvaefaipe 590

Qy 1001 NGWIAMNKEILIAAVSIGAIGMLIMVLMMLTMWRMEGPGKYNSDIKEMGGMHPIGTDTY 1060
Db 591 nswiamnkeilvaavsigaigmlimvlmmltmwrmegpgkynsdikemggmhpigtdty 650

Qy 1061 NSWILAKNLVLEAVEKVKFGAIEMSLMDMLKNEIKEVIETKHHKLESNHEIRDQMETVN 1120
Db 651 nswilaknlvleavekvkfgaiemslmdmlknevkevietkhhklesnheirdqmetvn 710

Qy 1121 AQPTHMQWRGIQPFETPPAAQIKCFTQSVSQSSHAKNSGKLTPKPKGPVLMMQSQPPSISP 1180
Db 711 aqpthmqwrgiqpfetppaaqikcftqsvsqsshaknsgkltpkpkgpvlmmqsqppsisp 770

Qy 1181 NKTINCRPCNIVGSEFSLVGMQWSPKIPKSGAKSYLIIDVCDKYFSTGIIIGKLIKTDA 1240
Db 771 nktincrpcnivgsefslvgmqwspkipksgaksylidvcdkyfstgiiigklikthda 830

Qy 1241 AGLKKAVMDMFERIYMTQYSRFKPTIVTDQGSEFVNNTVAGWCKSLGIKQLFTSVAHPQ 1300
Db 831 aglkkavmdmferiymtqysrfkptivtdqgsefvnntvagwcksLgikqlftsvahpq 890

Qy 1301 ANGQVEVSNRELKKILEKTVSSSRKDWVLKLPFVLLGLRTSPKEDLGFSPAELVYGSPL 1360
Db 891 angqvevsnrelkkilektvssskdwvlklpfvllglrtspkedlgfspaelvygspl 950

Qy 1361 RLGEFVAVWEGLEDEPEEWIKPVTLEHFYRLRDTLEDTWRRARENLQKAQDRQKYYDRK 1420
Db 951 rlgefvavwegledepeewikpvtlehfyrlrdtledtwrrarenlqkaqdrqkyydrk 1010

Qy 1421 GNWSHPSATQGHLGARLQRFRQPVMAGQEVSEALCDHLQKPLAGTEPCNIRDPARWTTS 1480
Db 1011 gnwshpsatqghlgarlqrfrqpvmagqevsealcdhlqkplagtepcnirdeparwtts 1070

Qy 1481 VWSQVGVVCGELYHSRLVTCELEKANGTVLVVSPFACAFHKRPLCEKPCTPHPVLWEPG 1540
Db 1071 vwsqvgvvcgelyhsrlvtcelekangtvlvvspfacafhkrplcekpctphpvlwepg 1130

Qy 1541 NRCPDNCMCRAVRMQQRHTACQHNSCLSNCPTHRFTLRPNTIGCACDVCWHPCTCWKPCT 1600
Db 1131 nrcpqrcmcrqvrmqqrhtacqhnsCdsncddrkrptlrnctstacdvcwhtcpwkpct 1190

Qy 1601 HLNLCSLDRYKQRCCQSCQEG 1621
Db 1191 hlnlcsldrykqrccqscqeq 1211

RESULT 5
AAE10608
ID AAE10608 standard; Protein; 1281 AA.
XX
AC AAE10608;
XX
DT 10-DEC-2001 (first entry)
XX
DE Human [illegible] protein, NOV-1.
XX
KW Human; NOV X; [illegible] protein; [illegible] protein; tumour;
KW trypsin inhibitor-like protein; gene therapy; haematopoietic; illness;
KW immunological disorder; neurodegenerative disorder; Alzheimer's disease;
KW Parkinson's disease; immunomodulatory; pharmacogenomic; haemostatic;
KW human immunodeficiency virus; HIV; fertility disorder; neuroprotective;
KW cytostatic; nootropic; anti infertility; cancer; chromosome 15;
KW NOV 1 protein.
XX
OS Homo sapiens.
XX
PN WO200162928 A2.
XX
PD 30-AUG-2001.
XX
PF 26-FEB-2001; 2001WO-US06151.
XX
PR 25-FEB-2000; 2000US-0184951.
PR 28-FEB-2000; 2000US-0185142.
PR 01-MAR-2000; 2000US-0185967.
PR 18-APR-2000; 2000US-0197724.
PR 27-APR-2000; 2000US-0199952.
PR 23-FEB-2001; 2001US-0789390.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Vernet CAM, Fernandes E, Shimkets RA, Macdougall J, Spaderna SK;
XX
DR WPI; 2001-582051/65.
DR N-PSDB; AAD17756.
XX
PT New isolated [illegible] like, [illegible] like, [illegible] like
PT polypeptide for diagnosing and treating pathological disorders, such as
PT Parkinson's disease, and for use in pharmacogenomics -
XX
PS Claim 1; Page 7; 189pp; English.
XX
CC The invention relates to novel human polypeptides referred as NOV X
CC and their corresponding nucleic acid sequences. NOV X collectively
CC includes NOV 1, NOV 2a and NOV 2b which are novel PIAA1223 like
CC polypeptides, NOV 3a, NOV 3b, NOV 3c and NOV 3d which are novel
CC STE20-like polypeptides and NOV 4a, NOV 4b, NOV 4c, NOV 4d and NOV 4e
CC which are novel trypsin inhibitor-like polypeptides. NOV X is used to
CC identify a potential therapeutic agent that can modulate its activity
CC and can be used for treating a pathology related to aberrant expression

```
CC   or aberrant physiological interactions of NOV X. NOV X or its DNA is
CC   used to determine the presence or predisposition to a disease associated
CC   with altered levels of NOV X, NOV X DNA and its mutants. NOV X or
CC   fragments thereof may be used to vaccinate animals to elicit an immune
CC   response specific for NOV X. NOV X or fragments thereof may be useful
CC   to treat or prevent a pathology associated with NOV X. The pathologic/
CC   states that can be treated or prevented are haematopoietic, cancer,
CC   immunological, tumour, neurodegenerative (e.g. Alzheimer's and
CC   Parkinson's disease), human immunodeficiency virus (HIV), kidney and
CC   fertility disorders. NOV X and its DNA are used in pharmacogenomics for
CC   predictive medicine. NOV X DNA is used in gene therapy. The present
CC   sequence is human novel KIAA1277 like protein. NOV 1, NOV 1 gene is
CC   located at chromosome 15.
XX
SQ   Sequence  1281 AA;
```

```
  Query Match             45.5%;  Score 2080;  DB 22;  Length 1281;
  Best Local Similarity   99.0%;  Pred. No. 0;
  Matches 1270;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy     411 MPYDHEVQFLPKWEHNFWTACSVSA AAYQRKSEVPVEESMHGELLQVEEWKCMYAFKEKV 470
           ||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db       1 mpydhtqplprwehnpwtacsvsa aayqrksevpveesmhgellqveewkcmyafkekv  60

Qy     471 MQICNELFDCPEWIAMEWSQCTIVTCGKLRYRVVLCTNHHCLHVGQCNFQLKCHIKEEVI 530
           |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db      61 mqtcnelfdcpewiamewsqctivtcgklryrvvlctnhhclhvgqcnfqlkchikeevi 120

Qy     531 PIPCYKPKEKSGPVEAKLPWLKQAQTLEEICIATEEPIEIPEPWSACSLDCSFGVQVKEVK 590
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 pipcykpkekspveaklpwlkqaqtleeiciateepieipepwsacsldcsfgvqvkevk 180

Qy     591 CRVLLPTQTETELPEEECTGPKPTERPCLLEACDESPASKELQHTLEGSELIYDWEY 648
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 crvllptqtetelpeeectgpkptcrpclleacdespasrelqhtlegselivdwey 238

Qy     649 AGETPCIATCLGGHQEAIAVCLHCPQQIVNESLCRMVHEPPAMSLACNIFPCPPGWHVS 708
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     239 agetpctatclgghqeaiavclhcpqqivneslcrmvheppamslacnifpcppgwhvs 298

Qy     709 SWQPCSATCGVGIQHRVVTCIRSHDAPPEECPDPKPHALQACNLGPCPIDGWHTEPWPEE 768
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     299 swqpcsatcgvgiqhrvvtcirshdappeecpdpkphalqacnlgpcpidgwhtepwpee 358

Qy     769 CSRTCGGGIQNRRVICVQLLDGSTLNLSAELCQGPKASSHKSCAKIPPPLEAVSWSK 826
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     359 csrtcggglqnrrvicvqlldgstlnlsaelcqgpkasshkscakippplevswsk 416

Qy     827 PSVSSCVVCTQVPKQVVQELAAKYPHIEIVIEQTGSSCVKKAGSRNRSKRAVHIKDVEM 886
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     417 psvssccvctqvpkqvvqelaakypheievieqtgsscvkkagsrnrskravhikdvem 476

Qy     887 EDAQDYLVQRVYTQIREEKETNETDSPAYLLPNISVLEKCIVPEEQKCTQWEKCC 944
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     477 edaqdylvqrvytqireeketnetdspaylllpnisvlekcivpeeqkctqwekcc 534

Qy     945 KCSLNRPRFSHLIKSGSLKCHCTAAPLCVYSCIASGAQEIVVLKTLTCNPLIAPAGFE 1004
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     535 kcslnrprfshliksgslkchctaaplcvysciasgaqeivvlktltcnplmapagfe 594

Qy    1005 PMPFYLQMLHGGEANSLCVIWHKMLDWNNKNCLYLGKGHICNLDPTLAIGMCNGASCP 1064
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     595 pmpfylqmlhggeanslcviwhkmldwnnknclylgkghicnldptlaigmcngascp 654

Qy    1065 NSWPTPNPCIAAVKLVAYSMSLAPTEIPPMCLDMSVDSFAGLIVAGTLAK 1114
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     655 nswedkvkptcivklpvgsmslapteiippmcldmsvdsfaglivagtlak 704

Qy    1115 AQPTHMQWRGTQEETPPAAGLEGEGVSQVSHAKNPCKLPPKPCTEVLMPQSQPPCTSP 1174
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     705 aqpthmqwrgtqeetppaaglegegvsqvshaknpckLppkpctevlmpqsqppctsp 764

Qy    1175 NKTINSPTSNIVYTIKPTEVINITCPLTPSEAIYIWTKDPCTLAQTSVKTLTLAGTGKTQT 1234
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     765 nktinsptsnivytikptevinitcpltpseaiyiwtkdpctlaqtsvktltlagtgktqt 824

Db     825 NKTINSPTSNIVYTIKPTEVINITCPLTPSEAIYIWTKDPCTLAQTSVKTLTLAGTGKTQT 884

Qy    1235 HENLQSLDKYKQRQCQSQPQT 1291
           |||||||||||||||||||||
Db     885 hnnlqsldykqrqcqsqpqt 907

RESULT   6
AAB83320
ID   AAB83320 standard; protein; 926 AA.
XX
AC   AAB83320;
XX
DT   29-MAR-2002  (first entry)
XX
DE   NOV14 protein sequence.
XX
KW   NOV; cytostatic; contraceptive; anti-inflammatory; immunomodulatory;
KW   cardiovascular; casein kinase II phosphorylation site; contraception;
KW   serine/threonine kinase; Pontz-Jeghers syndrome; cellular proliferation;
KW   epidermal growth factor; cell development; apoptosis; cell adhesion;
KW   growth inhibition; cell structure; motility; cancer; immune disorder;
KW   inflammatory disorder; cellular adhesion disorder; heart attack;
KW   cardiovascular disease; hypertrophic cardiomyopathy; marfan syndrome;
KW   therapy; NOV1; NOV2; NOV3; NOV4; NOV5; NOV6; NOV7; NOV8; NOV9; NOV10;
KW   NOV11; NOV12; NOV13; NOV14; NOV15; NOV16.
XX
OS   Unidentified.
XX
PN   WO200166608-A2.
XX
PD   27-MAY-2001.
XX
PF   17-NOV-2000; 2000WO-US31545.
XX
PR   19-NOV-1999;  99US-166306P.
PR   29-NOV-1999;  99US-167985P.
PR   09-MAR-2000;  2000US-187844P.
PR   16-NOV-2000;  2000US-0715417.
XX
PA   (CURA ) CURAGEN CORP.
XX
PI   Shimkets RA, Lichenstein H, Vernet C, Fernandes E;
XX
DR   WPI; 2001-648194/74.
```

```
DR   N-PSDB: AAF87125.
XX
PT   Novel human polypeptides and the nucleic acids that encode them useful
PT   for preventing, diagnosing and treating e.g. cancer, inflammation and
PT   immune disorders
XX
PS   Claim 1; Page 43-45; 141pp; English.
XX
CC   This sequence is the NOVI4 protein. The invention relates to
CC   the NOVI NOVI6 proteins, and their coding sequences. The proteins have
CC   Cytostatic; contraceptive; antiinflammatory; immunomodulatory; and
CC   cardiovascular activities. The sequences may be used in the prevention,
CC   diagnosis and treatment of diseases associated with inappropriate NOVX
CC   expression. They may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of protein by expressing inactive proteins or to
CC   supplement the patients own production of protein. They are used to
CC   produce NOVX proteins, by inserting the nucleic acid into a cell and
CC   culturing it to express the protein. The DNA may be used as DNA probes in
CC   assays to detect and quantitate the presence of similar DNAs in samples,
CC   and which patients may need restorative therapy. The NOVX protein may
CC   also be used as antigens in the production of antibodies (Abs) against
CC   NOVX and in assays to identify modulators of NOVX expression and
CC   activity. The anti-NOVX Abs and antagonist are used to down regulate
CC   expression and activity. The anti-NOVX Abs are used for detecting the
CC   presence of NOVX in samples. Disorders that may be prevented, diagnosed
CC   and/or treated vary depending on the NOVX protein. NOVI, NOV3, NOV5,
CC   NOV7, NOV9 11 and NOV13 16 have casein kinase II (cK II) phosphorylation sites
CC   characteristic of serine/threonine kinases, and are used to treat
CC   kinase related disorders (e.g. Fragile X syndrome, cystic fibrosis,
CC   proliferation and contraception). NOV2 3, NOV6 and NOV8 are homologous to
CC   the epidermal growth factor (EGF)-like super family and are involved in,
CC   e.g. regulation of cell development, apoptosis, cell adhesion, growth
CC   migration, cell structure and motility and protein management, and are
CC   used to treat cancers, inflammatory disorders, immune disorders and
CC   cellular adhesion disorders. NOV6-10 are homologous to EGF-like
CC   fibrillin proteins and are used to treat cardiovascular disease e.g.
CC   hypertrophic cardiomyopathy, long-QT syndrome and marfan syndrome.
XX
SQ   Sequence   926 AA;
```

```
  Query Match          54.4%;  Score 5086;  DB 22;  Length 926;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches  925;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     414 MADFIPEIEYIAAEIVVVIVVVIIIILWSIIFIIYKYRTQIKQFLNMDPPGDPNSNCYASLVTLSSGTEDSDYINANYIKGVNEPREFIATQGPLPNTIADFWRMVWEQKSATIVMMTKLEEMGRVKCDQYWPADSEEMFYGNIIVQLTSHALLGKTVTIRTFKIRNIQTGELKLLQGPWREEIAVPLEFAIQKLGGPDRFMNTVCQHQRVEVLHCISGAGRSGTFCLVDICLLLMDKRKDPSSVDIKKVLLEMRKFRMGLIQTADQLRFSYLAVIEGAKFIMGDSSVQDQWKELSHEDLEPPPEHIPPPPRPPKRILEPHNGKCKEFFPNFQHVKSMEGNPWTKKEINATDKPRRQNSKAKNPIENSFNDVPSPASLEQLKAEKSRVHSNAELGSTANLTNVSELRSNRLRQGGISQEDFETLKMQGISDVMRRLQKQPMPVKKGAGAQRKATQLPVAVSRFSSKGSLQSIDSAEKASLLPAFEDKDSYVHVLEERENNLVLTMPSTETDMNTVNKAIAGETCRAKTFIATQGPLSHTIKDFWQMIFNYKVIVIVMVTELKERKEECAQYWPSEEEVLYQNIVVTMVSSFVEHEAWHIREFLLSKNCIGDSTSITHGHYQVRTETPETHSISVIDWHFGVPSCPESVKSFLNFIFTVRESGSLSQEHGPIVVHCSAGIGRTGTFIVIDAMLDMMHSERKVDVYGFVSRIRAQRCQMVQTDMQYVFIYQALLEHYLYGDTELEVTSLETHLQKIYNKIPGVSKMEKNSLGAKWDFNSADIYYDHNIKLFMPTPDQDKLNGFMQAVKQNDVFSKSGPNAVHIHCSAGAGRTGVLIACDVAYLLMKHDKIIKEFSILSIVRDQRNYMVQTEAQYVFIHDTLVEAIKSKTEVEVIDFIDDGHMAENSVGSSITHSYQERMVNLSTFKSPLDAKRAYIKTLPYNRSRVILQPIEGVKGSDYINANFIDGYRQKNAYIATQGPLPETIGDFWQMVFEWKSCSIVMLTELEERGQEKCAQYWPSDGMVSYGDITVELKKEKECEGYTVRDLLVTNTRENKSRQIRQFHFHGWPEVGVPDDGKGMISIIAAVQKQQQQSGNHPITVHCSAGAGRTGTFCALSTVLERVKAEGILDVFQTVKSLRLQRPHMVQTEDQYQLCYRAALEYLGSFDHYASVEEKTKVDFDKTKLKDQIKSSQHNSRVSKIRPQNAFVATSALSEVNELREQNESLRAELLAVRGSQPSRSSYYSVLTDLQSRDSVDDSKLSACDFEFPRNYFVENADEFDRPAQKYIKQEMPPPLNQFLLFDNQRSTVVAALDTRVKLHAVLEDGSDGSPQFIRGAFASFHSKQWKKIHLSQDFKITLRCKIGEKNRYFNLMGKDKPISRSHDADTINPLKETIHIHNHQENKVRSLENDLMKEKEKMSKTIQLAFLQTQLDNAKCLISDLEQAKLRLEDEVKGRTAELEEFKSLNQGVIKDFQSKNIRFGSFVKKSFPDTEDQAKFTDDQKPLKLVIAVKTYLDEIKDLNITWQRDNLAAETGLTVSYKPAFSWLDPDYNPPVEGMKLDQDSLIWDYKREGFSIIMVTKQDEDTMGKKDISGDIQVLPYNDHRQLKDKLTKVITHPHFEWFVAKTPKVLSGLGTKLLGTAETNDDYQQVEKQMHYLAGLIMSEGYISATAKQSEICFTNKDKEFLDDFKHLLESLGFNFKIKEYDRNSYNIIISGQLVENYLDLIGPCSHKSKDKFVPEIFKYGNRQVRLAFLRGLFDTDGSVTIPRIVKSNASIKSGYQLRLQIYTNFRPQLQVLQKMLNTLGIENSIHIQKSNCCRLVVSRKDSLKLFMNIIGFGIEHKQNKLNIWAKAYNVTSRPGKNDPDPVLKEWKRVTGLNRKKAVDGIEKEAEMCPGFVTDVYDLEVEDNHNFFANGILVHNSYFGFTSDDTLDSVKEAYQGWTGTTIRDALENSITPAQFQTFLSMHSLQGFCNVVFDVWGALNPASSVPSLEHHHHHH
Db       1 madlipeieyiaaeivvviviiilwsiifiiykyrtqikqflnmdppgdpnsncyaslvtlssgtedsdyinanyikgvnepreflatqgplpntiadfwrmvweqksativmmtkleemgrvkcdqywpadseemfygniivqltshallgktvtirtfkirniqtgelkllqgpwreeiavplefaiqklggpdrfmntvcqhqrvevlhcisgagrsgtfclvdiclllmdkrkdpssvdikkvllemrkfrmglIqtadqlrfsylaviegakfimgdssvqdqwkelshedlepppehipppprppkrilephngkckeffpnfqhvksmegnpwtkkeinatdkprrqnskaknpienSfndvpspaslEqlkaekSrvhsnaelgstanltnvselrsnrlrqggisqedfetlkmqgisdvmrrlqkqpmpvkkgaGaqrkatqlpvavsrfssKgslqsidsaekaslLpafedkdsyvhvleerennlvltmpstetdmntvnkaiagetcrakTfiatqgplshtikdfwqmifnykvivivmvtelkerkeecaqywpseegvlyqnivvtmvssfvehEawhirefllsknCigdstsithghyqvrtetpethsisvidwhfgvpscpesvksflnfiftvresgslsqehgpivvhcsagigrtgtfividamldmmhserkvdvygfvsriraqrcqmvqtdmqyvfiyqallehylygdtelevtsletHlqkiynkipgvskmeknslgakwdfnsadiyydhnikLfmptpdqdklngfmqavkqndvfsksgpnavhihcsagagrtgvliacdvaylLmkhdkiikefsilsivrdqrnymvqtEaqyvfihdtlveaikskteveviDfiddghmaensvgssithsyqermvnlstfkspldakrayiktlpynrsrvilqpiegvkgsdyinanfidgyrqknayiatqgplpetigdfwqmvfewkscsivmltelEergqekcaqywpsdgmvsygditvelkkekecegytvrdllvtntrenksrqirqfhfhgwpevgvpddgkgmisiiaavqkqqqqsgnhpitvhcsagagrtgtfcalstvlervkaegildvfqtvkslrlqrphmvqtedqyqlcyraaleylgsfdhyasvEekTkvdfdktklkdqiksSqhnsrvskirpqnafvatsalsevnelrEqnesLraelLavrgsqpsrssyysvltdlqsrdsvddsklsacdfefprnyfvenadefdrpaqkyikqempppLnqfllfdnqrstvvaaldtrvklhavledgsdgspqfirgafasfhskqwkkihlsqdfkitlrckigeknryfnlmgkdkpisrshdadtinplketihihnhqenkvrslendlmkekekmsktiqlaflqtqldnakclisdleqaklrledevkgrtaeleefkslnqgvikdfqsknirfgsfvkksfpdtedqakftddqkplklviavktyldeikdlnitwqrdnlaaetgltvsykpafswldpdynppvegmkldqdsliwdykregfsiimvtkqdedtmgkkdisgdiqvlpyndhrqlkdklTkvithphfewfvaktpkvlsglgtkllgtaetndDyqqvekqmhylaglimseGyisatakqseicftnkdkeflddfkhllesLgfnfkikeydrnsynIiisgqlvenylDligpcshkskdkfvpeifkygnrqvrlafLrglfdtdgsvtiprivksnasiksgyqlrlqiytnfrpqlqvlqkmlntlgienSihiqksnccrlvvsrkdslklfmniigfgiehkqnklniwakaynvtsrpgkndpdpvlkewkrvtglnrkkavdgiekeaemcpgfvtdvydlevedNhnffangilvhnsyfgftsddtldsvkeayqgwtgttirdalensitpaqfqtflsmhslqgfcnvvfdvwgalnpassvpslehhhhhh
Db     414 MADFIPEIEYIAAEIVVVIVVVIIIILWSIIFIIYKYRTQIKQFLNMDPPGDPNSNCYASLVTLSSGTEDSDYINANYIKGVNEPREFIATQGPLPNTIADFWRMVWEQKSATIVMMTKLEEMGRVKCDQYWPADSEEMFYGNIIVQLTSHALLGKTVTIRTFKIRNIQTGELKLLQGPWREEIAVPLEFAIQKLGGPDRFMNTVCQHQRVEVLHCISGAGRSGTFCLVDICLLLMDKRKDPSSVDIKKVLLEMRKFRMGLIQTADQLRFSYLAVIEGAKFIMGDSSVQDQWKELSHEDLEPPPEHIPPPPRPPKRILEPHNGKCKEFFPNFQHVKSMEGNPWTKKEINATDKPRRQNSKAKNPIENSFNDVPSPASLEQLKAEKSRVHSNAELGSTANLTNVSELRSNRLRQGGISQEDFETLKMQGISDVMRRLQKQPMPVKKGAGAQRKATQLPVAVSRFSSKGSLQSIDSAEKASLLPAFEDKDSYVHVLEERENNLVLTMPSTETDMNTVNKAIAGETCRAKTFIATQGPLSHTIKDFWQMIFNYKVIVIVMVTELKERKEECAQYWPSEEEVLYQNIVVTMVSSFVEHEAWHIREFLLSKNCIGDSTSITHGHYQVRTETPETHSISVIDWHFGVPSCPESVKSFLNFIFTVRESGSLSQEHGPIVVHCSAGIGRTGTFIVIDAMLDMMHSERKVDVYGFVSRIRAQRCQMVQTDMQYVFIYQALLEHYLYGDTELEVTSLETHLQKIYNKIPGVSKMEKNSLGAKWDFNSADIYYDHNIKLFMPTPDQDKLNGFMQAVKQNDVFSKSGPNAVHIHCSAGAGRTGVLIACDVAYLLMKHDKIIKEFSILSIVRDQRNYMVQTEAQYVFIHDTLVEAIKSKTEVEVIDFIDDGHMAENSVGSSITHSYQERMVNLSTFKSPLDAKRAYIKTLPYNRSRVILQPIEGVKGSDYINANFIDGYRQKNAYIATQGPLPETIGDFWQMVFEWKSCSIVMLTELEERGQEKCAQYWPSDGMVSYGDITVELKKEKECEGYTVRDLLVTNTRENKSRQIRQFHFHGWPEVGVPDDGKGMISIIAAVQKQQQQSGNHPITVHCSAGAGRTGTFCALSTVLERVKAEGILDVFQTVKSLRLQRPHMVQTEDQYQLCYRAALEYLGSFDHYASVEEKTKVDFDKTKLKDQIKSSQHNSRVSKIRPQNAFVATSALSEVNELREQNESLRAELLAVRGSQPSRSSYYSVLTDLQSRDSVDDSKLSACDFEFPRNYFVENADEFDRPAQKYIKQEMPPPLNQFLLFDNQRSTVVAALDTRVKLHAVLEDGSDGSPQFIRGAFASFHSKQWKKIHLSQDFKITLRCKIGEKNRYFNLMGKDKPISRSHDADTINPLKETIHIHNHQENKVRSLENDLMKEKEKMSKTIQLAFLQTQLDNAKCLISDLEQAKLRLEDEVKGRTAELEEFKSLNQGVIKDFQSKNIRFGSFVKKSFPDTEDQAKFTDDQKPLKLVIAVKTYLDEIKDLNITWQRDNLAAETGLTVSYKPAFSWLDPDYNPPVEGMKLDQDSLIWDYKREGFSIIMVTKQDEDTMGKKDISGDIQVLPYNDHRQLKDKLTKVITHPHFEWFVAKTPKVLSGLGTKLLGTAETNDDYQQVEKQMHYLAGLIMSEGYISATAKQSEICFTNKDKEFLDDFKHLLESLGFNFKIKEYDRNSYNIIISGQLVENYLDLIGPCSHKSKDKFVPEIFKYGNRQVRLAFLRGLFDTDGSVTIPRIVKSNASIKSGYQLRLQIYTNFRPQLQVLQKMLNTLGIENSIHIQKSNCCRLVVSRKDSLKLFMNIIGFGIEHKQNKLNIWAKAYNVTSRPGKNDPDPVLKEWKRVTGLNRKKAVDGIEKEAEMCPGFVTDVYDLEVEDNHNFFANGILVHNSYFGFTSDDTLDSVKEAYQGWTGTTIRDALENSITPAQFQTFLSMHSLQGFCNVVFDVWGALNPASSVPSLEHHHHHH 926

Qy     734 GETPAPPEECRDEKPHALQACNQFDCPPCPWHIFFWQQCSRTCGKGTQNRRVTCRQLITDG 793
Db     422 qetpappeecrdekphalqacnqfdcppgwhiffwqqcsrtcgkgtqnrrvtcrqlitdg 480

Qy     794 SLLNGSKLLQLRRAGMPYMDLAVNLHSHVKDIESKFPSEMKFIEGQEYEIVDTIVEGEEEP 853
Db     481 stlnlsdelcqapkaeehkcmrtdpyphlrqfwkmvwehksvscvvmlttqlekgrekcaqy 540

Qy     854 RPIPSSEMMCPELPCFDLVPDSQMDPCSKIPSEMKFIEGQEYEIVDTIVEGEEEPRPIPSS 913
Db     541 rrlplsemmcrdlpafplcrsapppcskipsemkfiegqeyeivdtivegeeeprpipss 600

Qy     914 INITIGSPAYLIPNIPDVIKCDVEPPQKSICWFKWQECSCQNCKFEYMTDFSVSKIKFHCLA 973
Db     601 inltiqsrayllpnipdvikcdvepfqksicwfkwqecscqnckfeymtdfsvskikfhcla 660

Qy     974 APDIDVYRCIAGSAQEIVVLKLIGTPNKLIARPALREPMREYPGMDHSEANSLGVTWHKM 1033
Db     661 apdidvyrciagsaqeivvlkligtpnkliarpalrepmreypgmdhseansLgvtwhkm 720

Qy    1034 RQMWNNKNDIYLTDDDDTSNQPFLRALTIGHCFSNSAGSINSWRIKNKQFEAAVKQGAYSMDT 1093
Db     721 rqmwnnknditlddddtsnqpflralTIghcfsnsagsinswrikNkqfeaavkqgaysmdt 780

Qy    1094 AQFTFLIPNMSQLMPTCFVSDPLASQIIYQLVAELAKAQPTHMQWPGIQEFTPPAAQLPG 1153
Db     781 aqftflipnmsqlmptcfvsdplasqiiyqlvaelakaqpthmqwpgiqeftppaaqlpg 840

Qy    1154 EIQAVGLGDAMSKDSIPGERTMVSLVTNWTAESGDKGILPEQLTSLIKTEVINS 1208
Db     841 etqsvgsgshakasqklqfkpkqrrlargsqpgsisinktlnsriqntvyitktevini 900

Qy    1209 IGMLPPDSPAVYGWPEVGPLQPVP 1232
Db     901 lgdlitpseatytwtkdqtllqpvpk 926


RESULT   7
AAB40500
ID   AAB40500 standard; Protein; 886 AA.
XX
AC   AAB40500;
XX
DT   08-FEB-2001  (first entry)
XX
DE   Human  NOVX  protein polypeptide sequence SEQ ID NO:428.
XX
KW   Human, open reading frame, ORFX, detection, cytostatic, hepatotropic;
KW   vulnerary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW   anticonvulsant, osteopathic, antiarthritic, immunosuppressant; cardiant;
KW   immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW   hypotensive, dermatologic, immunodeficiency, antiinflammatory;
KW   antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW   antianaemic, gene therapy, cancer, proliferative disorder; hypertension;
KW   neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW   cardiovascular disease, diabetes mellitus, hypothyroidism; SCID; AIDS;
KW   cholesterol ester storage; systemic lupus erythematosus; infection;
KW   extracorporeal immunodeficiency, anaemia, autoimmune disorder, asthma;
KW   allergy, aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW   bone damage, cartilage damage, antiinflammatory disease, coagulation;
KW   thrombosis; contraceptive.
XX
OS   Homo sapiens.
XX
PN   WO200058473-A2.
XX
PD   05-OCT-2000.
XX
PF   31-MAR-2000; 2000WO-US08621.
XX
PR   31-MAR-1999;   99US-0127607.
PR   02-APR-1999;   99US-0127636.
```

```
PR   05 APR 1999;  99US-08127706.
PR   30 MAR 2000;  2000US-00540763
XX
PA   (CURA ) CURAGEN CORP.
XX
PI   Shimkets RA,  Lennon M;
XX
DR   WPI; 2000-572652/52.
DR   N-PSDB; AAC74446.
XX
PT   Novel nucleic acids and polypeptides derived from open reading frames X
PT   useful for treating cell, cancers, proliferative disorders,
PT   neurodegenerative disorders and cardiovascular disease
XX
PS   Claim 11; Page 698-699; 760pp; English.
XX
CC   AAC74446 to AAC77606 encode the proteins given in AAB14287 to AAB14967,
CC   which represent the human ORFX open reading frames 1 to 8161. The ORFX
CC   sequences have activities such as: cytostatic; hepatotropic; vulnerary;
CC   antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
CC   osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC   immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC   antidiabetic; hypotensive; dermatological; immunosuppressive;
CC   antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC   antithyroid; and antianaemic. The sequences can be used for determining
CC   the presence of or predisposition to, or preventing or treating
CC   pathological conditions associated with an ORFX-associated disorder. The
CC   nucleic acids can be used to express ORFX proteins in gene therapy
CC   vectors. The proteins and nucleic acids may be used to treat cancers,
CC   proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC   graft vs host disease, cardiovascular disease, diabetes mellitus,
CC   hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC   erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC   bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC   allergies, apoptotic anaemias, burns, wounds, bone and cartilage damage,
CC   nocturnal haemoglobinuria, anti-inflammatory diseases, endocrine
CC   modulation, to inhibit thrombosis, and as a contraceptive.
XX
SQ   Sequence  889 AA;
```

Query Match 50.2%; Score 4693; DB 21; Length 889;
Best Local Similarity 99.2%; Pred. No. 0.2e-300;
Matches 876; Conservative 0; Mismatches 5; Indels 2; Gaps 1;

```
Qy     591 CRVELLFTQTEEELPEDEYCFKSITEGCILEAVTESPASEELLELTFEVSELTRWRE 750
           ||||||||||||||||||||||.|:|||||| |||||||||||||||||| |:|| |:| |
Db       2 CRVILTTTQTETELPEEEECEQPKIPTERPETLEADESPASTRELDIPLPEDSETAYDEY 181

Qy     651 AGFTPCTATCTGGGHQEATAVCLHTQTQQTVNDSLCDMVIRFFAMSQACNTEFCTFKWHV 710
           |||||||||||||||||||||||:|||||||||||| ||||||||||||||||||||||||
Db      62 AQTTPETATCTQQTHEPANAVETHTQTPQTTNDSTSTDMVHRPPAMSQACNTEFCTFKWHV 121

Qy     711 SWGFCSATSVGVGTQTRGVYCLRPSETFAEPEECKDEKPHALQACNQFTCFHGWHTSEWQVC 770
           ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db     122 SWQPCSATQVGVGTQTRDVYCLHPDETPAPPEECRDEKPHALTACNQTDPPPAHCCWQFG 181

Qy     771 CSRTCGSGIQNRRVCNSLDNSMADGACVPEARPVEMEACVLPECPPDQFRSEETSCFMLV 830
           ||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db     182 CSRTCGSGIQNRRVCMSLDNSMADGACVPEARPVEMEACVLPECPPDQFRSEETSCFMLV 241

Qy     831 CSVSCGVCTQPPEQVCLSGANETVTLRMWQTQDSLCGVVPRDPLVEHLQGHMVEELTFLV 890
           |||||||||||||||||||| ||||||| |||||| |||||||||||||||||||||||||
Db     242 CSVSCGVCTQPPEQVCLPGANETVTLRMWQTQDSLCGVVPRDPLVEHLQGHMVEELTFLV 301

Qy     891 LGEQGPQVTLSVQVRVYLQTPFFKVRVLEGSAVLLPNITVTTKCTVPEEPCKLTLWEKLS 950
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     302 LGEQGPQVTLSVQVRVYLQTPFFKVRVLEGSAVLLPNITVTTKCTVPEEPCKLTLWEKLS 361

Qy     951 PCTQAGSAQEIVVELKTGTTAAEDTLVFFCTARGSAQEIVVELKTGTTAAEDTLVFFCTA 1010
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     362 PCTQAGSAQEIVVELKTGTTAAEDTLVFFCTARGSAQEIVVELKTGTTAAEDTLVFFCTA 421

Qy    1011 PMKEYPCMQRSEANSLSVTWHKMPCMWNNKNCLYLSSHHCNQPFLSATLSNCSACGSI 1070
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     422 PMKEYPCMQRSEANSLSVTWHKMPCMWNNKNCLYLSSHHCNQPFLSATLSNCSACGSI 481

Qy    1071 NCWEEKNKCPSAAVKCTAYSMCIAETELTWNMSCLMLSGHVSGDLANQLIYQLVAELAK 1130
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     482 NCWEEKNKCPSAAVKCTAYSMCIAETELTWNMSCLMLSGHVSGDLANQLIYQLVAELAK 541

Qy    1131 ACPTHMCWRCTAQEEHPAAQTRCPSVNQSSHAKNSGKLTLKPKKQPVSMKQCSVHPSISE 1190
           |||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db     542 ACPTHMCWRCTAQEEHPAAQTRCPSVSQSSHAKNSGKLTLKPKKQPISMKQCSVHPSISE 601

Qy    1191 NKTINSKISNIVYTIKKTEVINTCPDTTPSEATYTWTKDETTLQPSVKTTLDTTKTQT 1250
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     602 NKTINSKISNIVYTIKKTEVINTCPDTTPSEATYTWTKDETTLQPSVKTTLDTTKTQT 661

Qy    1251 QNPTRKEQASTYEPSVANHLGSLVEISSVELYAFAPVTLSVIERNTTKFFHNHLSVVVSLVE 1310
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     662 QNPTRKEQASTYEPSVANHLGSLVEISSVELYAFAPVTLSVIERNTTKFFHNHLSVVVSLVE 721

Qy    1311 AALSANVIAMETPVFSSVPQGNSTWRLASYYKGDSIIVFCVDSLNPVSLENSGIYVCIAI 1370
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     722 AALSANVIAMETPVFSSVPQGNSTWRLASYYKGDSIIVFCVDSLNPVSLENSGIYVCIAI 781

Qy    1371 NAIGKAVAISVELIDLEDKHWYFSKIVFLVGYKLYLGATNTKINSNLKDIGTFFFVIFFWEP 1430
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     782 NAIGKAVAISVELIDLEDKHWYFSKIVFLVGYKLYLGATNTKINSNLKDIGTFFFVIFFWEP 841

Qy    1431 GNWSPREAIGWLCILAVDGVLFLQLVMANDAVSEALCPDVLQKPLA 1475
           ||||||||||||||||||||| |||||||||||  | | |
Db     842 GNWSPRSATQHLSAIIQRPQPEQVMANDPEVSEALCPDVLQKPPEA 886
```

RESULT 8
AAE07808
ID AAE07808 standard; protein; 845 AA.
XX
AC AAE07808;
XX
DT 01-NOV-2001 (first entry)
XX
DE Novel human protein (NHP) #6.
XX
KW Novel human protein; NHP; thrombospondin; gene therapy; cancer;
KW cytostatic; antisense therapy; antiangiogenesis; biological disorder.
XX
OS Homo sapiens.
XX
PN WO200164011-A2.
XX
PD 23-AUG-2001.
XX
PF 15-FEB-2001; 2001WO-US05290.
XX
PR 17-FEB-2000; 2000US-0183282.
XX
PA (LEXI) LEXICON GENETICS INC.
XX
PI Donoho G, Scoville L, Turner A, Friedrich G, Zambrowicz B;
PI Sands AT;
XX
DR WPI; 2001-514226/56.
DR N-PSDB; AAD14899.
XX
PT An isolated nucleic acid encoding a thrombospondin useful as a
PT hybridization probe and gene therapy treatments of cancer
XX
PS Disclosure; Page 47-48; 50pp; English.
XX
CC The invention relates to novel human polynucleotides encoding proteins
CC that share sequence similarity with animal proteins having thrombospondin
CC repeats. NHPs are expressed in human cell lines, pituitary, lymph nodes,
CC prostate, testis, adrenal gland, uterus, foetal kidney, foetal lung and
```
```

```
CC   gene trapped human cells. NHPs are thrombospondins useful for treating
CC   biological disorders involving angiogenesis, cancer and development and
CC   also in pharmacogenomic applications. NHPs are useful as a hybridisation
CC   probe for screening libraries, assessing gene expression patterns and
CC   also in gene therapy. Proteins having thrombospondin repeats act as
CC   receptors, secreted extracellular matrix proteins and proteases. The
CC   present sequence is novel human protein (NHP).
XX
SQ   Sequence   845 AA;


  Query Match           50.2%;  Score 4689;  DB 22;  Length 845;
  Best Local Similarity  100.0%;  Pred. No. 1.6e-299;
  Matches  839;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MASWTSPWWVLIGMVFMHSPLPQTTAEKSPGAYFLPEFALSPQGSFLEDTTGFQFLTYRY 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 maswtspwwvligmvfmhsplpqttaekspgayflpefalspqgsfledttgfqfltyry 60

Qy      61 DDQTSRNTRSDFDKDGNWDAWGDWSDCSRTCGGGASYSLRRCLTGRNCFGQNIRYKTCSN 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ddqtsrntrsdfdkdgnwdawgdwsdcsrtcgggasyslrrclTgrncfgqniryktcsn 120

Qy     121 HDCPPDAEDFRAQGCSAYNDVQYQGHYYFWLPRYNDPAAPCALKCHAQGQNLVVELAPKV 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 hdcppdaedfraqgcsayndvqyqghyyfwlprYndpaapcalkchaqgqnlvvelapkv 180

Qy     181 LDGTPCNTDSLDMCISGICQAVCDDPLGSNAKFDNCGVCAGDGSTCRLVRGQSKSHVSP 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 ldgtprntdsldmcisgicqavcddplgsnakfdncgvcagdgstcrlvrgqskshvsp 240

Qy     241 EKREENVIAVPLGSRSVRITVKGPAHLTIFSKTLQGSKGFHSFNSPGVFVVENTTVTFQR 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 ekreenviavplgsrsvritvkgpahltifskTlqgskgfhsfnspgvfvventTvtfqr 300

Qy     301 GSFRQIFKTPGHIMADFIKTRYTAAPDSVVQFEYQPISHQWEQITPPPCIVIQDGGYV 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 gsfrqifktpghimadfiktryTaapdsvvqfeyqpishqweqitpppcivIqdggyv 360

Qy     361 LNSAECVDIPLFFVVPTHYCHYYPENVKPKPKLKSCSMDPCPSSLCEFEIMEYHMHFQPLP 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 lnsaecvdipLffvvpthychyypenvkpkpklkscsmdpcpssLcefeimeyhmhfqplp 420

Qy     421 EWFHNFWIACSVFTCGVLFSFFVCVLESMHCRIEQVEKWSVFHFHMVAHIRETVRQLTNEIK 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 ewfhnfwiacsvftcgvlfsffvcvlesmhcrieqvekwsvfhfhmvahiretvrqltneik 480

Qy     481 KWIAMEWSQDTVTQRQLRYRVVLEINHRGEHVQGENPQLKLHIKEEGNIPPEYKPKEK 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 kwiamewsqdtvtqrqlryrvvleinhrgehvqgenpqlklhikeegnipPeykpkek 540

Qy     541 SPVEAKLPWLKQAQELEETRIATEEPTTIPEPWSAQSTTCDPCVQVREVKCRVLLIPTQI 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 spveaklpwlkqaqeleetriateepttipepwsaqsttcdpcvqvrevkcrvlliptqi 600

Qy     601 ETELPEEECEGPKLPTERPCLLEACDESPASRELDIPLPEDSETTYDWEYAGFTPCTATC 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 etelpeeecegpklpterpcllEacdespasreldiplpedsettydweyagftpctatc 660

Qy     661 LGGHQEAIAVCLHIQQQIVNQSLCDMVHKPPAMSQACNIEICHPMWHVGSWGPCSAICG 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 lgghqeaiavclhiqqqivnqslcdmvhkppamsqacnieichpmwhvgswgpcsaicg 720

Qy     721 VGIQTRDVYCLHPGETPAPPEECKDEKPHALQACNQFDCPPGWHIEEWQQCSRTCGGGTQ 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 vgiqtrdvyclhpgetpappeecrdekphalqacnqfdcppgwhieewqqcsrtcgggtq 780

Qy     781 NRRVTCHQITTRSPENESHFTCQGPKASSHKSCARTPPPHLAVGFWSS 830
           |||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 nrrvtcrqlttdspttlalsdclqqppkrsckscrtdpplslrqlws 830

RESULT 6
```

```
AAE07869
ID   AAE07869 standard; Protein; 771 AA
XX
AC   AAE07869;
XX
DT   01-NOV-2001  (first entry)
XX
DE   Novel human protein (NHP) #7.
XX
KW   Novel human protein, NHP, thrombospondin; gene therapy; cancer;
KW   cytostatic, antisense therapy, angiogenesis, biological disorder.
XX
OS   Homo sapiens.
XX
PN   WO200161011-A2.
XX
PD   23-AUG-2001.
XX
PF   15-FEB-2001; 2001WO-US05290.
XX
PP   17-FEB-2000; 2000US-0183282
XX
PA   (LEXI-) LEXICON GENETICS INC.
XX
PI   Donoho G,  Scoville J,  Turner CA,  Friedrich G,  Zambrowicz B;
PI   Sands AT;
XX
DR   WPI; 2001-514776/56.
DR   N-PSDB; AAD14370.
XX
PT   An isolated nucleic acid encoding a thrombospondin useful as a
PT   hybridization probe and gene therapy treatments of cancer -
XX
PS   Disclosure; Page 46-48; 56pp; English.
XX
CC   The invention relates to novel human polynucleotides encoding proteins
CC   that share sequence similarity with animal proteins having thrombospondin
CC   repeats. NHPs are expressed in human cell lines, pituitary, lymph nodes,
CC   prostrate, testis, adrenal gland, uterus, foetal kidney, foetal lung and
CC   gene trapped human cells. NHPs are thrombospondins useful for treating
CC   biological disorders involving angiogenesis, cancer and development and
CC   also in pharmacogenomic applications. NHPs are useful as a hybridisation
CC   probe for screening libraries, assessing gene expression patterns and
CC   also in gene therapy. Proteins having thrombospondin repeats act as
CC   receptors, secreted extracellular matrix proteins and proteases. The
CC   present sequence is novel human protein (NHP).
XX
SQ   Sequence   771 AA;


  Query Match           45.2%;  Score 4225;  DB 22;  Length 771;
  Best Local Similarity  91.1%;  Pred. No. 4.5e-269;
  Matches  756;  Conservative    0,  Mismatches    0,  Indels   74;  Gaps    1;

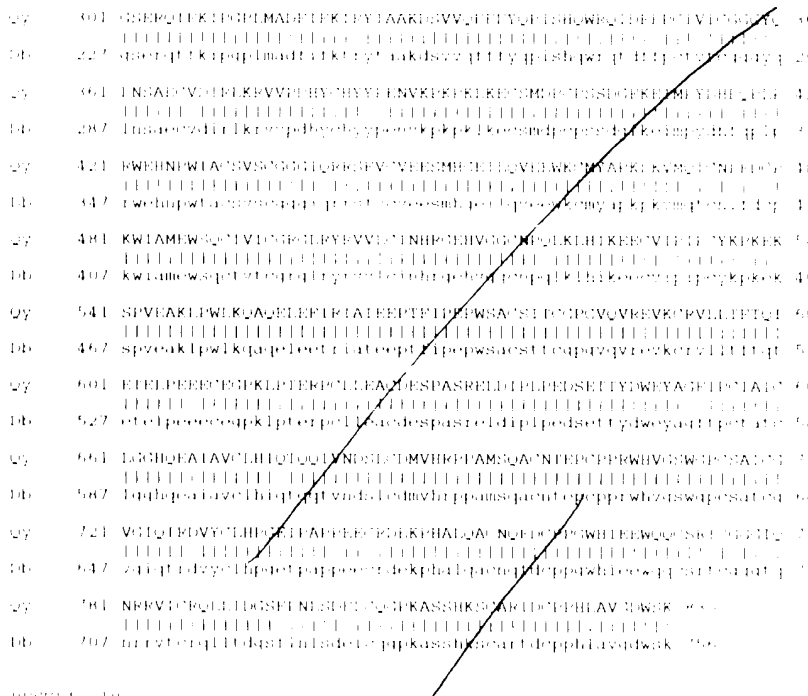
Qy       1 MASWTSPWWVLIGMVFMHSPLPQTTAEKSPGAYFLPEFALSPQGSFLEDTTGFQFLTYRY 60
           ||||||||||||||||||||||||||||||
Db       1 maswtspwwvligmvfmhsplpqttacksp---------------------------- 30

Qy      61 DDQTSRNTRSDFDKDGNWDAWGDWSDCSRTCGGGASYSLRRCLTGRNCFGQNIRYKTCSN 120
                                                 ||||||||||||||||
Db      31 ----------------------------------------grnceqqnirykTcsn 46

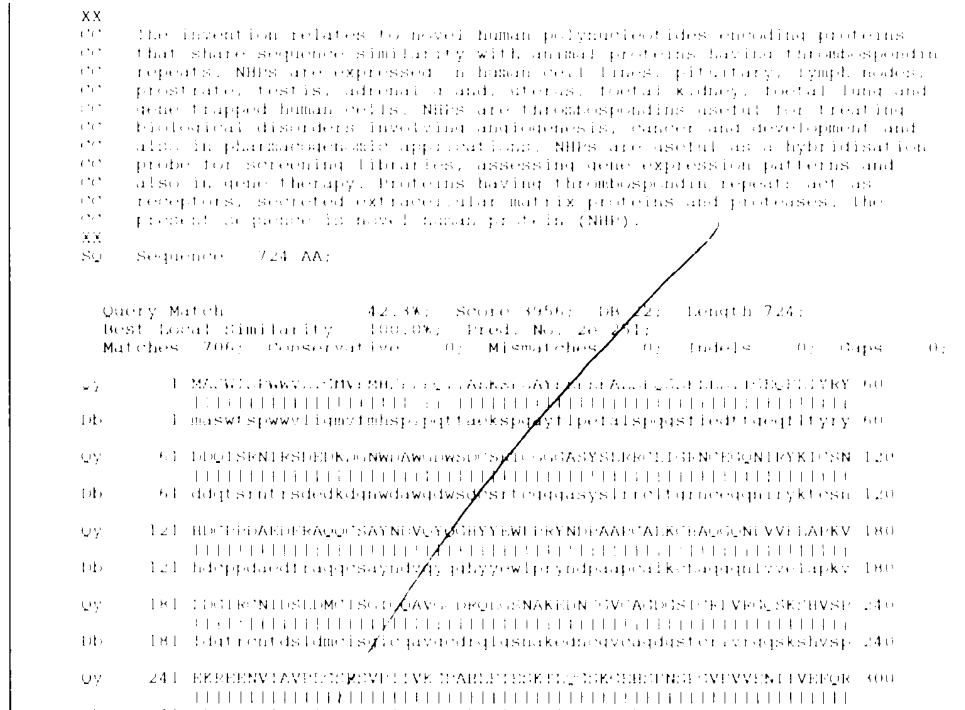
Qy     121 HDCPPDAEDFRAQGCSAYNDVQYQGHYYFWLPRYNDPAAPCALKCHAQGQNLVVELAPKV 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      47 hdcppdaedfraqgcsayndvqyqghyyfwlprYndpaapcalkchaqgqnlvvelapkv 106

Qy     181 LDGTPCNTDSLDMCISGICQAVCDDPLGSNAKFDNCGVCAGDGSTCRLVRGQSKSHVSP 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     107 ldgtprntdsldmcisgicqavcddplgsnakfdncgvcagdgstcrlvrgqskshvsp 166

Qy     241 EKREENVIAVPLGSRSVRITVKGPAHLTIFSKTLQGSKGFHSFNSPGVFVVENTTVTFQR 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     167 ekreenviavplgsrsvritvkgpahltifskTlqgskgfhsfnspgvfvventTvtfqr 226
```



RESULT 10
AAE07866
ID AAE07866 standard; Protein 724 AA.
XX
AC AAE07866;
XX
DT 01-NOV-2001 (first entry)
XX
DE Novel human protein (NHP) #4.
XX
KW Novel human protein; NHP; thrombospondin; gene therapy; cancer;
KW cytostatic; antisense therapy; angiogenesis; biological disorder.
XX
OS Homo sapiens.
XX
PN WO200164011-A2.
XX
PD 23-AUG-2001
XX
PF 15-FEB-2001; 2001WO-US05203.
XX
PR 17-FEB-2000; 2000US-0183282.
XX
PA (LEXI) LEXICON GENETICS INC.
XX
PI Iomobo G; Scoville S; Turner CA; Friedrich G; Zambrowicz B;
PI Sands AT;
XX
DR WPI; 2001-514776/56.
DR N-PSDB; AAD14467.
XX
PT An isolated nucleic acid encoding a thrombospondin useful as a
PT hybridization probe and gene therapy treatments of cancer
XX
PS Disclosure; Page 48-49; 56pp; English.



XX
CC The invention relates to novel human polypeptides encoding proteins
CC that share sequence similarity with animal proteins having thrombospondin
CC repeats. NHPs are expressed in human cell lines, pituitary, lymph nodes,
CC prostate, testis, adrenal gland, uterus, foetal kidney, foetal liver and
CC gene trapped human cells. NHPs are thrombospondins useful for treating
CC biological disorders involving angiogenesis, cancer and development and
CC also in pharmacogenomic applications. NHPs are useful as a hybridisation
CC probe for screening libraries, assessing gene expression patterns and
CC also in gene therapy. Proteins having thrombospondin repeats act as
CC receptors, secreted extracellular matrix proteins and proteases. The
CC present sequence is novel human protein (NHP).
XX
SQ Sequence 724 AA;

RESULT 11
AAB99651
ID AAB99651 standard; Protein; 1745 AA.
XX

```
AC   AAB90551;
XX
DT   01-JUN-2001  (first entry)
XX
DE   Human secreted protein, SEQ ID NO: 86
XX
KW   Human; secreted protein; immunomodulatory, anti-inflammatory, anti HIV, cytostatic; cardiant;
KW   dermatological, antiinflammatory, anti HIV, cytostatic; cardiant;
KW   vascular, arteriosclerosis, pathological, neuropsychiatric,
KW   nootropic; anticonvulsant; antialzheimers; antiparkinsonian;
KW   antimicrobial; pulmonary; cytokine; gene therapy; cancer;
KW   protein coordinate data; infection.
XX
OS   Homo sapiens.
XX
PN   WO200121658-A1.
XX
PD   29-MAR-2001.
XX
PF   22-SEP-2000; 2000WO-US26013.
XX
PR   24-SEP-1999;   99US-0155709.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ni J, Baker KP, Birse CE, Ebner R, Fiscella M, Komatsoulis GA;
PI   Lafleur DW, Moore PA, Olsen HS, Rosen CA, Ruben SA, Soppet DR;
PI   Young PE, Wei P, Florence KA;
XX
DR   WPI; 2001-235311/24.
DR   N-PSDB; AAF97891.
XX
PT   Nucleic acids encoding 32 human secreted polypeptides, useful for
PT   preventing, diagnosing and/or treating e.g. cancers, Parkinson's
PT   disease and diabetic retinopathy.
XX
PS   Claim 11; Fig 4; 899pp; English.
XX
CC   The present sequence is one of 32 novel human secreted polypeptides. The
CC   nucleic acid molecules and polypeptides may be used in the prevention,
CC   diagnosis and treatment of diseases such as immune disorders (e.g.
CC   multiple sclerosis, systemic lupus erythematosus and human
CC   immuno-deficiency virus (HIV) infections), hyperproliferative disorders
CC   (e.g. cancers and Gaucher's disease), cardiovascular diseases
CC   (e.g. Scimitar syndrome, Chagas' cardiomyopathy and [illegible]
CC   arteriosclerosis), antiobesity disorders (e.g. corneal graft
CC   neovascularisation), neuropsychiatric disorders
CC   (e.g. Huntington's chorea, Alzheimer's disease and Parkinson's disease),
CC   infectious diseases and/or [illegible] wound healing, regeneration
CC   and/or chemotaxis. The nucleic acid molecules may be used to produce the
CC   secreted polypeptides. They may also be used as PCR primers for
CC   assays to detect and quantitate the presence of similar nucleic acid
CC   sequences in samples. The polypeptides may be used as antigens in the
CC   production of antibodies and in assays to identify modulators of
CC   their expression and activity.
XX
SQ   Sequence  1745 AA;
```

```
  Query Match             41.3%;  Score 3861;  DB 22;  Length 1745;
  Best Local Similarity   42.9%;  Pred. No. 1.2e-244;
  Matches  756;  Conservative  264;  Mismatches  552;  Indels  204;  Gaps   72;

Qy      64 TSRNTKSLEDKVCNWGAWCWSVCVFTCCCASYSLRECLTCRNCEECNNESYKECSNEWC 123
           ||| |||||||| |||| |||||||||||||||||| |||||||||||||| ||
Db      22 ssrtarsmedrddlwdaqqawswsvfvqqqasysirrclsskscqrqrqlyrrcsnvdc 81

Qy     124 PPDAEDFRAQQCSAYNGVLYGHYYEWLPRYNGPAAPCALKCHAQCQNLVVELAPKVLDG 183
           ||:| |||||||||:|:||| | ||||| ||| ||:||| |:| ||||||||||
Db      82 ppeaqdfraqqpsahndrkfhqqtyowlpvsndpdnpcslkcpakattlvvelapkvldg 141

Qy     184 TRCNTDSLLMKFLGLDAVTFRKLSGNAKFTDNVPVAGSKFHVSPEKR 230
           ||| ||||||||||:|| |||| |||| |||||||| ||||||||||| || :| |
Db     142 tpcytaslgmelsgleqlgpedhplqstvkedpqvgeqdgstgrlcrqgykspisatfq 201

Qy     231 LLNVVAVFLLSGVPLVRGFAHTLEYLAVDSLDNVESTFNSVFGVFVLNTFVEPGKGSE 403
           :: |:|:| ||| :|: :||| ||:|:|||||:||||:| :| | ||:|:|:|:|: :
Db     202 HPvvqlpggerhlcl IPqqdhdyhcfkkpqpkpvclssqtqflcfksscqgptpd 261

Qy     404 EQLTPEIPSFLVAAFTIPFSSVTAAEICVVLFFVQLFSVPLWEQFFFSVTSSSTLRS 463
           ::  :: ||| |||| | | : : || ||| |||| |:|||||||||: ||||||| |
Db     262 kvclimaqqlladflkkrnqeadcqgtcqqpqlcnrwretddffpcsatcvqqqyqlts 321

Qy     464 AECVVDLPLKFVVFDSVCVVFLNVKFFFKLKLCVMSDCFFSSICFFEIMPYTHFQPLPWF 523
           ||| |:|  ||| | |||||||||:|||||:||::|||||  |||
Db     322 aeqydlrsarrvsdqyqkqyppcqkpkplqecnldpcpa---------- rwc 364

Qy     524 HNFWIACDVSCQGIQSKSFVCVEESMGKEIQVEFNACMYAANSVMQIQNLFDCPKWI 580
             |||||| |||||| |:  |||| : | :  |||||||| ||  : | ||:|||||:
Db     365 atpwtanssscpqtiqcrrmcrcsfqphclcrcvkwmqtpkmpirqpcvifdpkwi 424

Qy     581 AMEWSQCIVLDGSLLKTKVVLQNHFGLDVSLKQLALLRSLCVLTIDVNSKHKSRV 637
           | ||| |||||||:|||||||||:||| | |||:|: | |||||||:| |||||||| ||
Db     425 aqewspctvtcqpqlryrmvlcidkqpcbtqrcspkkpkfkvcvicptpcypkpkpklpv 484

Qy     638 EAKLPWLKQAQELEETRIAIIHIPIPIPWSACGFTCGPGVQVREVKCRVLLIFTQTETE 697
           |||||| ||||||||    :|||:|||: |||||: ||| | ||| |:|:|||:|:|: :
Db     485 eakipwfkqaqqlvcqqaqrsqpqtipkrwsrtvtrqmtqprirvcpqellrdqprvad 544

Qy     698 LPEEECEGPKLPTERPCLIEEACDESPASKELDIPLP----------EDSETTYDWEYAGF 747
           || :|||||  ::|    ||  | : :|    |          :| :  |||| ||
Db     545 lpideceqpkpasqr-----acyaqpcsqcipcfnpdetdqlfqqlqdldelydweyeqt 599

Qy     748 TPCTATCLGGHQEAIAVCLHIQTQQTVNDSLCDMVHRPPAMSQACNIEPCPPRWHVGSWG 807
           | |: || ||| :|  |||:   :|||    ||| | :||| |||| || |
Db     600 tkcsrsqggpqdrvcvcvkqtcpqcqpvtlrqgqpflmvcctpdqcnfvcppqawpeq 659

Qy     808 PCSATCGVCIQTRDVYCLH-----PGETPAPPEE-CRDEKPHALQACNQFDCPPGWHIEE 861
           ||| ||||| |||||:| |        ||    :| ||  ||  :||||:|:||| |:  :
Db     660 pcstteqrqlqtrdvfcqhllrcqmactvlladerqpkpqtrqacnrfncppqawpeq 719

Qy     862 WQQCSRPCGGGTQNRFVTCMQLETHCGPLNESRPVQVCPKASSHKSCAWTWCTPDLAVCD 921
           || |||||||| | | | |:| : |||| | :  |  |  :: | :|| :  |||    : |
Db     720 wqpcsrtcrqgtqnrqvrcvkqtcpqcqplvtqggqpfdtvscsqcpdqkpsqtrqcnk 779

Qy     922 WSPCSVSCGVGIQFRQVVCQSLAAPCPPEIQPEMMCTGIPCPPDWIVEGQMPPVPFIPSDM 981
           |:|| ||| | | :|::  |  :: :|  ||    :| | :  |::
Db     780 wtscsasvgegtptrsaicrkaikthstcgnstiqpplpfsssirpcmlatcar-parp 838

Qy     982 PIPFCPCPCPIPVLSVSLPLPPAILGGPVLPLPCPVSPLPSLIQWE 1027
           ||     | | : :|||||| :::::  :|  |||| |:|:::|| || :| || ||
Db     839 pcs-----pqrlrsrcrqpqrqcpkfrs-rrtqlqpkrsclrpqss-rkplitwe 895

Qy    1028 KLPFCPCPCPIPVLSVSLPLPPAILGGPVLPLPCPVSPLPSLIQWEKLLARP 1080
           ||| | :|  :  :  | |||| |  | || | || |:|||  :|:|||
Db     896 kdqqhlissthvtvaptqyikihrlkpcdatcytcvapaprehfvikliqnrklvarpl 955

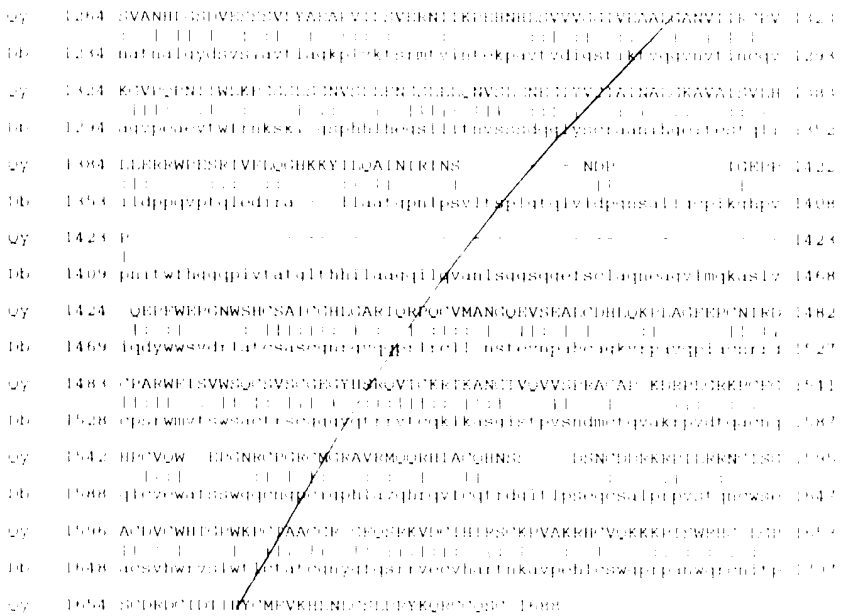
Qy    1081          AIEFPMPFVPCMLHPFANSLSVTWHP  MPCMWNNPNDLYLDDDHISNQPF 1055
                    | |:  :    |  | :    |  |  :  |   |  || :|
Db     956 sprsccecrlqrkpgpkvclpfkkqcqlsvskcrrqlarqfgsrqvcddllsr 1010

Qy    1056 LRALLGHCSNSACSTNSWELKNKQPFAAVKQ---CA--------YSMDPAQ--PDELIRN 1102
             ||      ||| ::    :  ||       ::| | |   |::: |
Db    1011    llcqqcwpgcellcswcaqdscrntticcedpqncqvillhlpfcmctcqprrlddilqc 1067

Qy    1103 MSQLMETGEVSDDLASQLIYQLVAELAKAQPIHMQ--WRGIQEETPP----------A 1148
           :|| |  |:|  :  |: ||  |    :  : | |          :
Db    1068 lcqqpc elrdlyskkhcaqlcpltlcsbtlcdcpdtllkpscerrtscpctlcpkkccqcls 1125

Qy    1149 AGIP    QETCSVSCCSHAKNCCKLTPKPCTPVLMPC SCPPSTCPNYTINSKRIGNIVV 1203
           : ||    |: | |:  | |           | ::|: |    :| :: : : :| ||
Db    1126 ssirtsstqdaqqqsrrpkrk----------ptilrkisaqqqlsasccvv-thlqqtva 1174

Qy    1204 ::EEIEVIKIIDLLIIIcLA...W.KL..:..:.:SVS:...SIGKI:.:NLI:EEQGIYEC 1263
           :  | : : |: |   | :| ::| :|:|    :|  | |  : | | |
Db    1175 lasqtlsvllhceaiqhprptlswarnqeewqtsdrillqpddslqllapweadvqtyec 1234
```



ID ABB10246 standard; Protein; 1766 AA.
AC ABB10246;
DT 16-JAN-2002 (first entry)
DE Human cDNA SEQ ID NO: 554.
KW Human; gene therapy; neural disorder; immune system disorder;
KW muscular disorder; reproductive disorder; gastrointestinal disorder;
KW pulmonary disorder; cardiovascular disorder; renal disorder;
KW proliferative disorder; inflammation.
OS Homo sapiens.
PN WO200154474-A2.
PD 02-AUG-2001.
PF 17-JAN-2001; 2001WO-US01434.
PR 31-JAN-2000; 2000US-179222P.
PR 04-FEB-2000; 2000US-180626P.
PR 24-FEB-2000; 2000US-184504P.
PR 02-MAR-2000; 2000US-186590P.
PR 15-MAR-2000; 2000US-189087P.
PR 17-MAR-2000; 2000US-190076P.
PR 18-APR-2000; 2000US-198124P.
PR 10-MAY-2000; 2000US-205515P.
PR 07-JUN-2000; 2000US-209465P.
PR 28-JUN-2000; 2000US-214886P.
PR 30-JUN-2000; 2000US-215345P.
PR 07-JUL-2000; 2000US-216647P.
PR 07-JUL-2000; 2000US-216880P.
PR 11-JUL-2000; 2000US-217486P.
PR 11-JUL-2000; 2000US-217496P.
PR 14-JUL-2000; 2000US-218290P.
PR 26-JUL-2000; 2000US-220964P.
PR 26-JUL-2000; 2000US-220964P.
PR 14-AUG-2000; 2000US-224519P.
PR 14-AUG-2000; 2000US-224519P.
PR 14-AUG-2000; 2000US-225215P.
PR 14-AUG-2000; 2000US-225214P.
PR 14-AUG-2000; 2000US-225266P.
PR 14-AUG-2000; 2000US-225267P.
PR 14-AUG-2000; 2000US-225268P.
PR 14-AUG-2000; 2000US-225270P.
PR 14-AUG-2000; 2000US-225447P.
PR 14-AUG-2000; 2000US-225752P.
PR 14-AUG-2000; 2000US-225758P.
PR 14-AUG-2000; 2000US-225759P.
PR 18-AUG-2000; 2000US-226279P.
PR 22-AUG-2000; 2000US-226681P.
PR 22-AUG-2000; 2000US-226868P.
PR 22-AUG-2000; 2000US-227182P.
PR 23-AUG-2000; 2000US-227000P.
PR 30-AUG-2000; 2000US-228924P.
PR 01-SEP-2000; 2000US-229283P.
PR 01-SEP-2000; 2000US-229443P.
PR 01-SEP-2000; 2000US-229444P.
PR 01-SEP-2000; 2000US-229445P.
PR 05-SEP-2000; 2000US-229566P.
PR 05-SEP-2000; 2000US-229713P.
PR 06-SEP-2000; 2000US-230483P.
PR 06-SEP-2000; 2000US-230489P.
PR 08-SEP-2000; 2000US-231243P.
PR 08-SEP-2000; 2000US-231243P.
PR 08-SEP-2000; 2000US-231244P.
PR 08-SEP-2000; 2000US-231413P.
PR 08-SEP-2000; 2000US-231414P.
PR 08-SEP-2000; 2000US-232080P.
PR 08-SEP-2000; 2000US-232081P.
PR 12-SEP-2000; 2000US-232968P.
PR 14-SEP-2000; 2000US-232397P.
PR 14-SEP-2000; 2000US-232398P.
PR 14-SEP-2000; 2000US-232399P.
PR 14-SEP-2000; 2000US-232400P.
PR 14-SEP-2000; 2000US-232401P.
PR 14-SEP-2000; 2000US-233005P.
PR 14-SEP-2000; 2000US-233064P.
PR 14-SEP-2000; 2000US-233065P.
PR 21-SEP-2000; 2000US-234227P.
PR 21-SEP-2000; 2000US-234274P.
PR 25-SEP-2000; 2000US-234997P.
PR 25-SEP-2000; 2000US-234998P.
PR 26-SEP-2000; 2000US-235084P.
PR 27-SEP-2000; 2000US-235834P.
PR 27-SEP-2000; 2000US-235836P.
PR 29-SEP-2000; 2000US-236327P.
PR 29-SEP-2000; 2000US-236667P.
PR 29-SEP-2000; 2000US-236668P.
PR 29-SEP-2000; 2000US-236669P.
PR 29-SEP-2000; 2000US-236670P.
PR 02-OCT-2000; 2000US-236892P.
PR 02-OCT-2000; 2000US-237037P.
PR 02-OCT-2000; 2000US-237038P.
PR 02-OCT-2000; 2000US-237039P.
PR 02-OCT-2000; 2000US-237040P.
PR 13-OCT-2000; 2000US-239935P.
PR 13-OCT-2000; 2000US-239937P.
PR 20-OCT-2000; 2000US-240966P.
PR 20-OCT-2000; 2000US-241221P.
PR 20-OCT-2000; 2000US-241785P.
PR 20-OCT-2000; 2000US-241786P.
PR 20-OCT-2000; 2000US-241787P.
PR 20-OCT-2000; 2000US-241808P.

PR 26-OCT-2000; 2000US-241809P.
PR 26-OCT-2000; 2000US-241824P.
PR 01-NOV-2000; 2000US-244617P.
PR 08-NOV-2000; 2000US-246474P.
PR 08-NOV-2000; 2000US-246475P.
PR 08-NOV-2000; 2000US-246476P.
PR 08-NOV-2000; 2000US-246477P.
PR 08-NOV-2000; 2000US-246478P.
PR 08-NOV-2000; 2000US-246524P.
PR 08-NOV-2000; 2000US-246525P.
PR 08-NOV-2000; 2000US-246526P.
PR 08-NOV-2000; 2000US-246527P.
PR 08-NOV-2000; 2000US-246528P.
PR 08-NOV-2000; 2000US-246529P.
PR 08-NOV-2000; 2000US-246530P.
PR 08-NOV-2000; 2000US-246610P.
PR 08-NOV-2000; 2000US-246611P.
PR 08-NOV-2000; 2000US-246612P.
PR 17-NOV-2000; 2000US-249207P.
PR 17-NOV-2000; 2000US-249208P.
PR 17-NOV-2000; 2000US-249209P.
PR 17-NOV-2000; 2000US-249210P.
PR 17-NOV-2000; 2000US-249211P.
PR 17-NOV-2000; 2000US-249212P.
PR 17-NOV-2000; 2000US-249213P.
PR 17-NOV-2000; 2000US-249214P.
PR 17-NOV-2000; 2000US-249215P.
PR 17-NOV-2000; 2000US-249216P.
PR 17-NOV-2000; 2000US-249217P.
PR 17-NOV-2000; 2000US-249218P.
PR 17-NOV-2000; 2000US-249244P.
PR 17-NOV-2000; 2000US-249245P.
PR 17-NOV-2000; 2000US-249263P.
PR 17-NOV-2000; 2000US-249265P.
PR 17-NOV-2000; 2000US-249267P.
PR 17-NOV-2000; 2000US-249269P.
PR 17-NOV-2000; 2000US-249270P.
PR 01-DEC-2000; 2000US-250160P.
PR 01-DEC-2000; 2000US-250204P.
PR 05-DEC-2000; 2000US-251030P.
PR 05-DEC-2000; 2000US-251988P.
PR 05-DEC-2000; 2000US-256719P.
PR 06-DEC-2000; 2000US-251479P.
PR 08-DEC-2000; 2000US-251858P.
PR 08-DEC-2000; 2000US-251862P.
PR 08-DEC-2000; 2000US-251869P.
PR 08-DEC-2000; 2000US-251889P.
PR 08-DEC-2000; 2000US-251990P.
PR 11-DEC-2000; 2000US-252497P.
PR 05-JAN-2001; 2001US-259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-476161/51.
DR N-PSDB; ABA96468.
XX
PI Isolated nucleic acid molecule encoding an inflammation-associated
PI polypeptide is used in preventing, treating or ameliorating a medical
PI condition -
XX
PS Claim 11; SEQ ID NO: 554; 859pp - Sequence Listing, English.
XX
CC The present invention provides human cDNAs, proteins and related genomic
CC DNAs. These can be used in the treatment of neural, immune system,
CC muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,
CC renal and proliferative disorders and inflammation. The present sequence
CC is a protein of the invention.
XX
SQ Sequence 1766 AA;

Query Match 41.3%; Score 2860.5; DB 22; Length 1766;
Best Local Similarity 42.1%; Pred. No. 1.4e-244;
Matches 765; Conservative 269; Mismatches 579; Indels 207; Gaps 33;

Qy 25 YARRSFQAYFILELALSLQGF LEGFLGLPLIYEYYLSQTSNIAGLILKNGWDAW 81
Db 1 tagechladysvsdsdsqkdpsmscprrartpatllllfatllssrtarseedrddqlwddaw 60

Qy 82 GDWSGCSRTYSSGAVSSLKSCLEGKNCLQNIKYKLCGNPSGTTEAFLEKAVQCSAYNEV 141
Db 61 qpwsscsrtysggavssikscleqknclqnikykicgnpsgttefflekavqcsaynev 120

Qy 142 QVQDPVYYFWLPQYMSPAAFCALPCHAGCQNLVVELAPEVLPCTPCNTPSLSMCIPCPQA 201
Db 121 khdwqlqswlpsvdpdqpcsk pdpcpl cdqpksdqtrcytcsidmcisqleqi 180

Qy 202 VGCDPQCGSNAVFLNCVACHCSTCWEPQFSFHVSFTKEEENVIAVPLGSRSVRTIV 261
Db 181 vqcsdhqlqstckedqvqlcnpdpstcdlcrqlksqlkpl nfk kp vipqrcdcir 240

Qy 262 KGVAGLIFIKRVQCNSLGNISCVLVVSLTLVLQSCSEGLFL TLRMAGDFE 321
Db 241 kgpdllqletkllqptkqcnslsatycfldnssdldqktqdkcllrmaqptlaqtiksi 300

Qy 322 FYTAAFCSVV LFFFYQPISHLWSCIFFPCTVTCGSLYQLNSAFTVICELFSVVPDHYCH 381
Db 301 rnsqsadstqtifqpfilswretdffpcsatcsaq qltsacqdisatcradqqk 360

Qy 382 YYIFNVFFFFTLFSCMCFCFCCCSPFIMEVIHFCLFSWERNFWIACSVSGGGTQKA 441
Db 361 yypcnikpkpklpcnnidpcpa tweafpwtacsssscqqdiqst 404

Qy 442 SLVCVLLMHCFILVLWEWCMTALFLFAVHQCNEFLCLRWCAHLWSQCLVCQNCLWKN 501
Db 405 acccddiqtlctacckwqqtpkqcqqcqiflqfalaqwaqatclqiqlcqt 463

Qy 502 VVLCINRGGEHVGCNFQLFLHIFLICVCFLFCYEIKEKCLVEAKLFWLFQAQLLEETRI 561
Db 464 vvlcidhrgdtqpwsqktkphikccciqptpcqkpkcklpcvakqfwkpaqclcqaa 523

Qy 562 ATFFFTFTFFFWSACSTSCGCVQVFVFCPVLIFTQTFTFLFFFCCCFLPTERECL 621
Db 524 ccccpclfpcwwacklqgcqcpclcc pltlafqcqdlpclbccqkpasqr--- 580

Qy 622 LLACDESPASKELQIPLP EDSLTTTDWEYAGFTPCTATCLGGHQFAIAVC 671
Db 581 arvaqpcsqclpcfqpdctdqllqqlqdlkdlydweycqttkcscccqqvcpqvvsc 638

Qy 672 LHILQTQQIVNGSLCLMVmklrAMSQACKtLrrrkNdVcqWGrcqArcGVcrQrkvvcc 731
Db 639 lnkqtrcpacenlwtsrcppqlfkscqdpcparwcdgkwspcstcgcqlqtrdvtcs 698

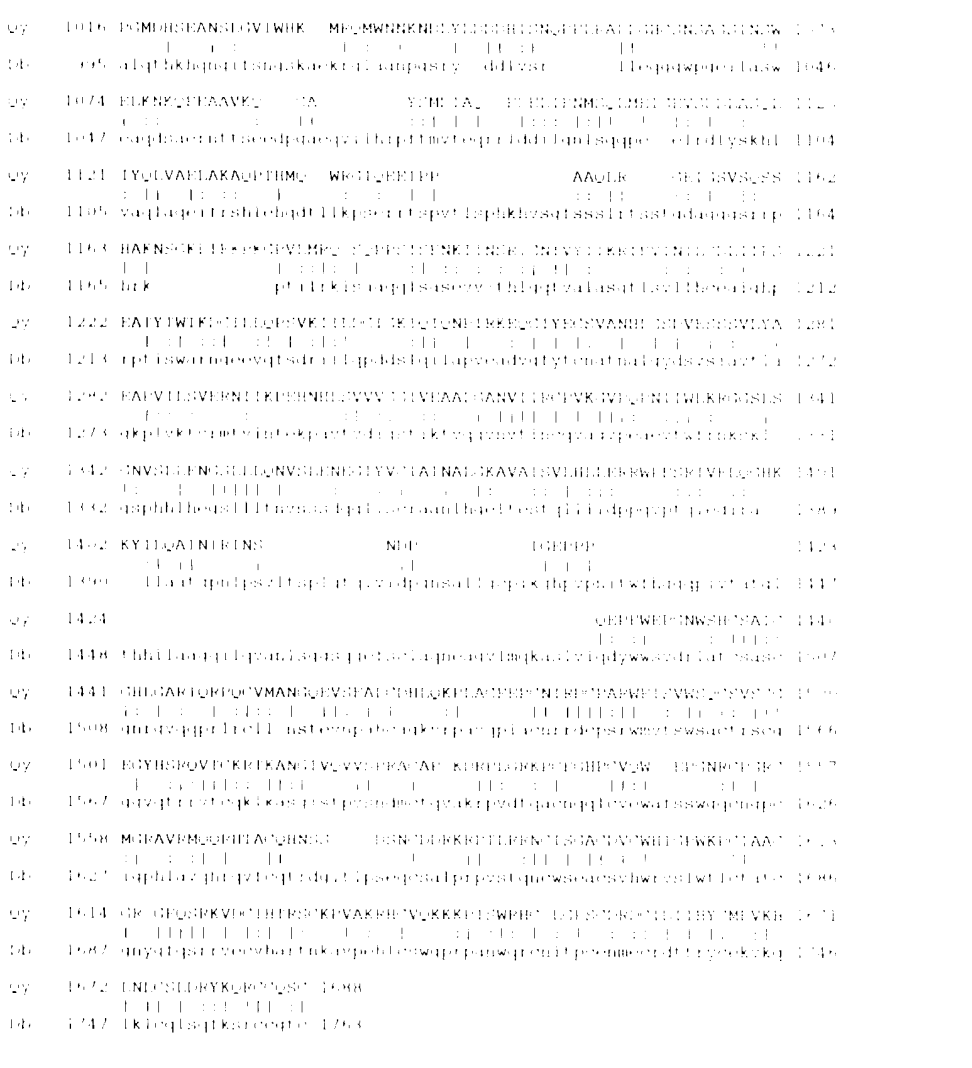
Qy 732 H-----PCETPAPPEE CEDEKPHALQACNQFTCPPGWHIDEWQQCGKTCSSTQNRKVI 785
Db 699 hllsremnetvllqdelcrqpkpstcqccrtncppawypnqwqpcstccqqrqkrovl 758

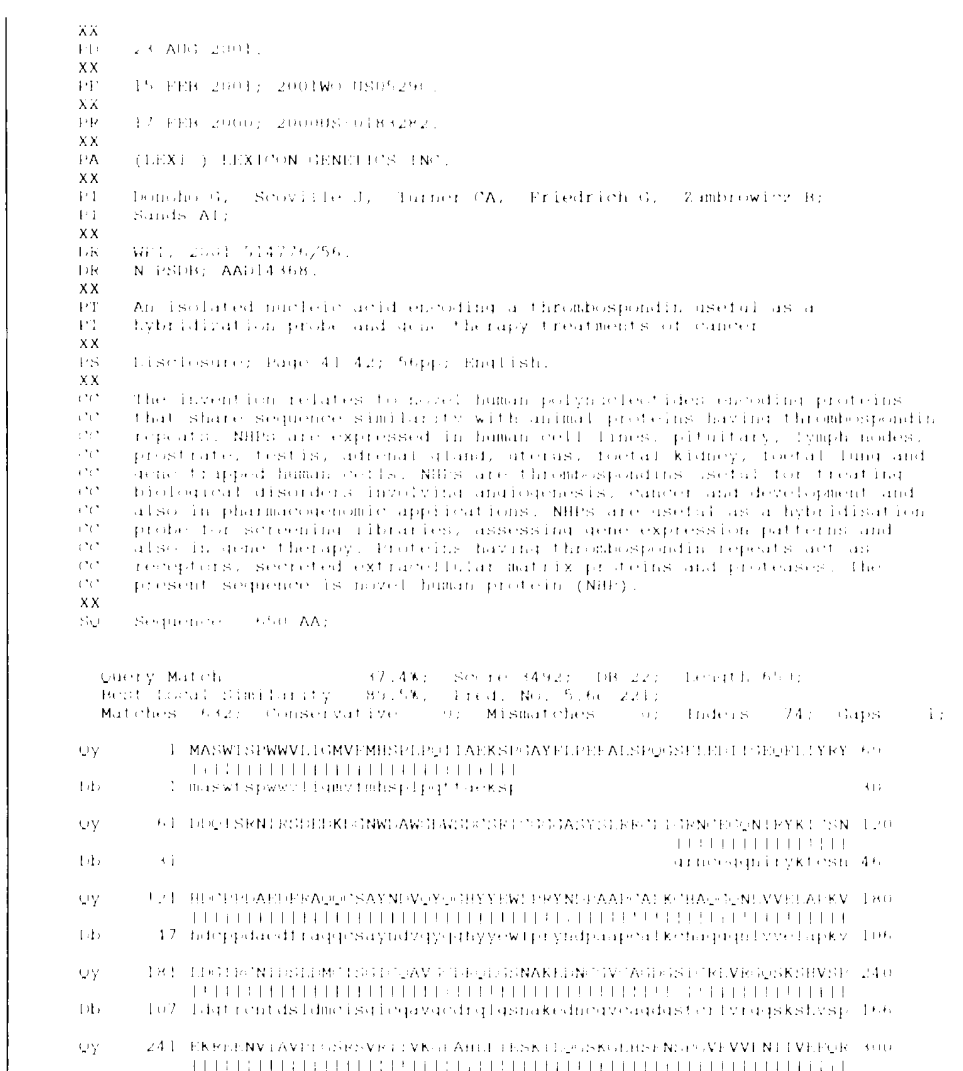
Qy 786 CRQLITCSFLNLSDELQQPKASSHKSCARIPCPHGLAVGHWSKCSVSCGVGYQKRSQV 845
Db 759 kprcqkqslelpctt sqskpqrqpckkdlpswllsdwtcstcqeqtqtrsal 818

Qy 846 CQPLAASCRRPLPSMMCMPCLPCPCVPSCQMPFCSPIESEMPIILQQQPQIISVQRVY 905
Db 819 crkmlkpqlstcmqscklqplfscrqcmctcqrqrpik kpkrrrkqy 874

Qy 906 IQTPEPPEINETTGFEAYLLNIPSVIECTPVEPQKSLQWEELQKTQQNLEELQTTESG 965
Db 875 lqtrrqrklhtcqptqqllqkca lspscrkpqtlwcklppblicclctvapiq 934

Qy 966 SLEIHCLAAPSICVYRCTACSAQEIVVLKLIGTLENKLIAKP----------ALKEPMKEY 1015
Db 935 yllikqlpcdrqcyccaqpcchlcklqaqklsatplprscccvlaqckpspkc 994





```
RESULT  14
AAE07867
ID   AAE07867 standard; Protein; 650 AA.
XX
AC   AAE07867;
XX
DT   31-MAY-2001  (first entry)
XX
DE   Novel human protein (NHP) #7.
XX
KW   Novel human protein; NHP; thrombospondin; gene therapy; cancer;
KW   cytostatic; anti-disease therapy; anti-angiogenesis; biological disorder.
XX
OS   Homo sapiens.
XX
PN   WO200107611-A2.
XX
PD   15-FEB-2001.
XX
PF   10-AUG-2000; 2000WO-US021882.
XX
PR   12-AUG-1999; 1999US-0148428P.
XX
PA   (LEXI ) LEXICON GENETICS INC.
XX
PI   Donoho G, Scoville J, Turner CA, Friedrich G, Zambrowicz B;
PI   Sands AT;
XX
DR   WPI; 2001-213276/56.
DR   N-PSDB; AAD14468.
XX
PT   An isolated nucleic acid encoding a thrombospondin useful as a
PT   hybridization probe and gene therapy treatments of cancer
XX
PS   Disclosure; Page 41-42; 56pp; English.
XX
CC   The invention relates to novel human polynucleotides encoding proteins
CC   that share sequence similarity with animal proteins having thrombospondin
CC   repeats. NHPs are expressed in human cell lines, pituitary, lymph nodes,
CC   prostate, testis, adrenal gland, uterus, foetal kidney, foetal lung and
CC   acute-stimulated human cells. NHPs are thrombospondins useful for treating
CC   biological disorders involving angiogenesis, cancer and development and
CC   also in pharmacogenomic applications. NHPs are useful as a hybridisation
CC   probe for screening libraries, assessing gene expression patterns and
CC   also in gene therapy. Proteins having thrombospondin repeats act as
CC   receptors, secreted extracellular matrix proteins and proteases. The
CC   present sequence is novel human protein (NHP).
XX
SQ   Sequence  650 AA;


  Query Match             47.4%;  Score 3492;  DB 22;  Length 650;
  Best Local Similarity   89.5%;  Pred. No. 5.6e-221;
  Matches  632;  Conservative   0;  Mismatches   0;  Indels   74;  Gaps   1;

Qy        1 MASWISPWWVLIGMVFMHSPLPGIIAEKSPGAYFLPPFEALSPQGSFTIEIIFKPLEYKY 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1 maswispwwvligmvfmhsplpgiiaekspgayflppfealspqgsftieiifkpleyky 60

Qy       61 DDQTSRNFRGEEKICNWDAWGRARSCSFTVSTGAQYCSKVEIKRNCTDQNTFYKLYSN 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       61 ddqtsrnfrgeekicnwdawgrarscsftvstgaqycskveikrnctdqntfykLysn 120

Qy      121 RLCPPLAEDFKAQLQSAYNEVQVQHIYYEWLPWYNLDAAPCAERCHAQCNLVVELAEKV 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      121 rlcppLaedfkaqLqsayNevQvqhiyyewLpwynLdaapcaercHaqcnLvvelaekv 180

Qy      181 DLSLNCNTISSLDMPSISEQLAVLTSLGSNAKELSNCTVDAGLSSLCKLVKSQLKSHVSP 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      181 dLsLNCNTISSLDMPSISEQLAVLTSLGSNAKELSNCTVDAGLSSLCKLVKSQLKSHVSP 240

Qy      241 EKPDLENVIAVFLQGPSVVKLVKSFAHLTESAIRGPSKGLDSIENCLPVEVVINLIVEFQF 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      241 eKpdLenviavflqgpsvvklvksfahLtesairgpskgLdsienclpvevvinLivefqf 300

Qy      301 GSERQTFKTIPGPLMADFTFKTRYTAAKDSVVQFFFYQPISHQWRQTDFFPCTVIDGGGYQ 360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      301 gserqtfktipgplmadftfktrytaakdsvvqfffyqpishqwrqtdffpctvidggGyq 360

Qy      361 LNGAELVVLKKEVVLDGSENLLRFEVSLLRMPSSLGFKFHPYFNIMPFYDHPQPLFP 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      361 lngaelvvlkkevvldgsenllrfevsllrmpssLgfkfhpyfnimpfydhpqplfp 420

Qy      421 RWFLNNFWVLVSVSVLSVSVLGSLIVFLLCMHKLQSVRIWKYMYAIPKFFVMLYGNLFDCP 480
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      421 rwflnnfwvlvsvsvlsvsvlgslivfllcmhklqsvriwkymyaipkffvmlygnlfdcp 480

Qy      481 KWIAMEWSLNIVIFYGKGLKYFVVLVTNHHSGEHVGGGMNLGLKLHTKEDKVIPEGCYKPKEK 540
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      481 kwiamewslnivifygkglkyfvvlvtnhhsgehvgggmnlglklhtkedkvipegcykpkek 540
```

```
Db    467 KWlamowsqctytcqrqlryrmclbdrqdcrrapqlblsfrrr clplpcypbsk 466

Qy    541 SPVEAKLPWLKQAQELEETRIAIEEPTFIPEPWSACSTICGPGVQVRKVKCRVLLTFTQT 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    467 spveaklpwlkqpqelsetrlstcqttspcpwsacstticgpgvqvrkckcrllttfgt 526

Qy    601 RTELPEEECEGPKLPTERPCLLEACDESPASKELDIPLPE-SETIYLWEYAGFIPCIAIC 659
          ||||||||||||||||||||||||||||||||||||||||| |||||||||||||||||||
Db    527 rtelpeeeceapklpterpcllearcdespaskeldiplpeesetiylweyagfipciaic 586

Qy    660 LGGHQEAIAVCLHICTQQIVNRSLCEMVHKFFAMSQACNTEEQFFR 706
          |||||||||||||||||||||||||||||||||||||||||||||||
Db    587 lggHqeaiavclhiqtqqivnrslcemvhkffamsqacntEEqffr 633

RESULT 14
AAB01431
ID   AAB01431 standard; Protein; 874 AA.
XX
AC   AAB01431;
XX
DT   20-OCT-2000 (first entry)
XX
DE   Human TANGO 224 (form 2).
XX
KW   TANGO; 128; 140; 197; 212; 213; 224; 239; modulating agent; asthma;
KW   graft versus host disease; rheumatoid arthritis; psoriasis;
KW   inflammatory bowel disease; septic shock; ulcerative colitis;
KW   Crohn's disease; chronic myelogenous leukemia; cancer; liver
KW   disease; Hodgkin's disease; osteoarthritis; Lyme's disease;
KW   cachexia; autoimmune disease; myasthemia gravis; autoimmune diabetes;
KW   systemic lupus erythematosus; transgenic animal; diagnosis;
KW   prognosis, prophylaxis, therapeutic, human.
XX
OS   Homo sapiens.
XX
PN   WO200079284 A1.
XX
PD   28-DEC-2000.
XX
PF   24-DEC-1999;  99WO-US31025
XX
PR   [illegible]-1998;  98US-0223546.
XX
PA   (MILL-) MILLENNIUM PHARM INC.
XX
PI   Holtzman DA;
XX
DR   WPI; 2000-149743/40.
DR   N-PSDB; AAA47482.
XX
PT   Novel nucleic acid sequences encoding TANGO-128, 140, 197, 212, 213,
PT   224 and 239 polypeptides useful for the treatment of asthma, rheumatoid
PT   arthritis, psoriasis and autoimmune diseases
XX
PS   Claim 8; Fig 30; 209pp; English.
XX
CC   Nucleic acids encoding TANGO polypeptides are useful as modulating
CC   agents for regulating cellular processes like asthma, graft
CC   versus-host diseases, rheumatoid arthritis, psoriasis, inflammatory
CC   bowel diseases, septic shock, ulcerative colitis, Crohn's diseases,
CC   chronic myelogenous leukemia, cancer, liver disease, Hodgkin's
CC   disease, osteoarthritis, Lyme's disease, cachexia and autoimmune
CC   diseases e.g. myasthemia gravis, autoimmune diabetes and systemic
CC   lupus erythematosus. The nucleic acids are also useful for producing
CC   transgenic animals and the TANGO polypeptides themselves. Partial
CC   TANGO-128, 140, 197, 212, 213, 224, 239 sequences are useful in
CC   forensic biology, for diagnostic assays, prognostic assays,
CC   pharmacogenomics and for monitoring clinical trials. TANGO
CC   polypeptides are suitable for both prophylactic and therapeutic
CC   methods for treating a subject at risk of a disorder or having a
CC   disorder associated with aberrant TANGO expression. A wide range
CC   of cellular disorders can be treated.
XX
SQ   Sequence  874 AA;

  Query Match          20.9%;  Score 2891;  DB 21;  Length 874;
  Best Local Similarity   58.4%;  Pred. No. 2.9e-181;
  Matches  443;  Conservative  126;  Mismatches  195;  Indels   30;  Gaps    5;

Qy      61 [illegible] 123
           :||  |||:|:|| ||||| |:|||||||||||||||||:: ::|:|:|||||:|||| ||
Db      22 sattqrseqdrdqlwdaweqpwsvsrtcqqsasysltclsskscsqrnirytctsnvdc 81

Qy     124 FFCVAFGFRALQSSAYNRVVYGGYYEWLFATNLFAAFCALSIIALQLRLVVELSAFEVLDG 183
           ||:| ||||||||:|||:: | :|||| :|| |||||| |:|  ||||||||||||
Db      82 ppcrqdfrqqpqrsabcfkklhgqfpcwlpvscqfdnpcqlkcqpakqtflcvelapkvldq 141

Qy     184 IFCMIGCFIMCISIFGLAVCTGLISNAFFVVSLVELVKFSVSFFSV 230
           ||| |:|||||||:|| |||| |||| |||||||| ||||||||||| || :|  |
Db     142 tcqytesllmclslcqlpcpedhqtlcfskfbcrpcn qflscvtcqrqpkcqlscfks 201

Qy     231 FFNVIAVPIGSRSVRITVKGPAHLFIESKILGSKGEHSFNSPGVFVVENITVFFQPQRGSE 303
           :: |:|:| ||| :|: :||| ||:::|||||:|||:| :| | |:|:|:::|:|: :
Db     202 dflcvaipygscbirlclkgpdblytctkflqptkqcnslsctatflcdnssedfqkfpd 261

Qy     304 FQTFKIPGFIMADFIFKTRYTAAKDSVVQFFFYQPISHQWPQTDFFPCTVTCGGGYQLNS 363
           ::  :: ||| |||| | | : : || ||| ||||| |:|||:|||||: |||||||| |
Db     262 kcilcmagpttadfickirnsgaadstvqfcfypplflhrwrtcdftpcsat.rqqlpts 321

Qy     364 AFCVDIRLKPVVFDRYCRYYEENVFEKFKLKECSMFPC---------PSSICFKEIMPYD 414
           ||| |:|  ||| | |||||||:||||||:|::|||         |:||||:|||||
Db     322 acvydlrsrrvvrdqycbyypccikpkplpcnlqlpcpaqrrlcrpnskayklmpyd 381

Qy     415 RFQFIPKWGRNLWIAVSVLCCILIGLSIVVLVLWSMGCRKMYAKNKSVMLDT 174
           : |||||  |||||| |||||| |:  |||| : | :  |||||||| ||  : | |
Db     382 lyhplprweatpwtacsssccqqqlqsrawscwedlqqhptsvcwwkcmytpkmplaqpc 441

Qy     475 NLLVSFPKWIAMDWSQCTVFCCILSYRVVLETNHRKGHHVGSCNPQLKLHIKEECVIPTPC 534
           |:|||||:| ||| |||||:|||||||||:|||| | |||:|: | |||||||:| ||
Db     442 nlfdcpkwlapwwqpf.cppllycc.lidlrrdrgqpsplpkcksc.cqtppc 501

Qy     535 YRIPKEESPVEASIPWFKQAQELEETRIAIEEPTFIPEPWSACSTICGPGVQVRKVKCRVL 594
           |||||| |||||||| |||||||    :|||:|||| |||: ||| | ||| |:|:||
Db     502 ykpkckklpwaklpwfkqaqelcqparscqqcplpcpswsacstticgpgvqvrkckcrll 561

Qy     595 LTFTQTETELPEEECEGPKLPTERPCLLEACDESPASKELDIPLP----------EDSET 644
           |:|:|:  :|| :|||||  :::|     ||  | | |:    |            :| :
Db     562 lsfsqsvadlpidccegpkpasqr      acyaqpcsqeipcfnpdctdqlfqqlqdtdc 616

Qy     645 TYPDWEYAGFTPCTATCLGGHQEAIAVCLHITQTQQTVNRSLCEMVHKFFAMSQACNTEEQFF 704
           ||||| ||| |: :| || |||:  ||: |||:   ::||    ||| : :|:| :||||
Db     617 lydwcycglfkccccpqgrpcc.cslkkgtqcqcrlcrtcrrpg glkcculdpcp 676

Qy     705 RWHVCSWGVCSATCGVSLQTRPVYCLH     EGETIAIPEF CRLEKPHALQAVNQFC 758
           || :| | ||| ||||:|||||:| |       ||   :| || ||  :||||:|:
Db     677 arwclqkwspcsltcgvqlqtrfvqclhcrcmctcfilkdclcrgpkpctcqacnrtlc 736

Qy     759 CPPCWHIEEWQQCSRTCGGGTQNRRVICRQLLIDGSPLNLSIELCQQPKASSHKSCARTD 818
           ||| |:  ||| |||||||| | | | |:| : ||||| | :  |  | :  :|| | |
Db     737 ypfcwtfdqwgqcsrtcggqtqnrrvicrqllidgspldlsiqlcqqpkasshksctrtd 796

Qy     819 CIPICAVCDWSEPKVSCVLPERRVQPAAPAGLTLVMVLVPIVVLVAVVAVLMP 872
           ||    : ||:|| ||| | | |  :|:::  |    ::  :|  ||   :| | :
Db     797 lpscwllcdwtcqtsqcqlqtcvlcrdlkldcldcmstlcppqlpdsssirpcmla 856

Qy     873 ECSK 882
           |:::
Db     857 tcar 860

RESULT 15
AAY87347
ID   AAY87347 standard; Protein; 643 AA.
XX
```

```
AC   AAY87432;
XX
DT   11 MAY 2000  (first entry)
XX
DE   Human signal peptide containing protein HSPP 124 SEQ ID NO:124.
XX
KW   Human; signal peptide containing protein; HSPP; diagnosis; cancer;
KW   inflammation; cardiovascular disease; antitumor; anti-inflammatory;
KW   antimicrobial; nootropic; neuroprotective; cardiovascular; hepatotropic;
KW   antiasthmatic; gene therapy; cell proliferation; neurological disorder;
KW   reproductive disorder; developmental disorder; arteriosclerosis;
KW   cirrhosis; psoriasis; acquired immune deficiency syndrome; anaemia;
KW   asthma; Crohn's disease; infection; Alzheimer's disease; schizophrenia;
KW   Parkinson's disease; Huntington's disease; ovulatory defect;
KW   muscular dystrophy.
XX
OS   Homo sapiens.
XX
PN   WO200006610-A2.
XX
PD   06-JAN-2000.
XX
PF   25-JUN-1999;  99WO-US014484.
XX
PR   26-JUN-1998;  98US-0090762.
PR   02-JUL-1998;  98US-0091264.
PR   06-OCT-1998;  98US-0103290.
PR   11-DEC-1998;  98US-0111729.
XX
PA   (INCY ) INCYTE PHARM INC.
XX
PI   Lal P;  Tang YT;  Corcoran JA;  Corley NC;  Gandhi K;  Swarnakar MF;
PI   Akerblom IE;  Au-Young J;  Yue H;  Patterson C;  Reddy R;  Hillman J;
PI   Bandman O;
XX
DR   WPI; 2000-144234/12.
DR   N-PSDB; AAZ08232.
XX
PT   New human signal peptide containing proteins useful in treatment,
PT   prevention and diagnosis of e.g. cancer, inflammation and
PT   cardiovascular disease.
XX
PS   Claim 1; Page 243; 427pp; English.
XX
CC   AAZ08109 to AAZ08242 encode AAY87224 to AAY87357 which represent the
CC   human signal peptide containing proteins HSPP-1 to HSPP-134. HSPPs have
CC   anticancer, anti-inflammatory, antimicrobial, nootropic, hepatotropic,
CC   neuroprotective, cardiovascular and antiasthmatic activities, and can
CC   be used in gene therapy. HSPPs can be used to treat or prevent disorders
CC   associated with decreased activity or function of HSPP. Antagonists of
CC   HSPP are used to treat or prevent disorders associated with increased
CC   activity or function of HSPP. Such diseases include cell proliferative
CC   (including cancer), inflammatory, cardiovascular, neurological,
CC   reproductive or developmental disorders, (e.g. arteriosclerosis,
CC   cirrhosis, psoriasis, acquired immune deficiency syndrome, anaemia,
CC   asthma, Crohn's disease, autoimmune or other infections, congestive or
CC   ischaemic heart disease, Alzheimer's, Parkinson's or Huntington's
CC   diseases, schizophrenia, ovulatory defects, muscular dystrophy). HSPP
CC   nucleic acids can be used for the recombinant production of HSPP, for
CC   detecting HSPP in standard hybridisation and amplification assays (e.g.
CC   diagnosis and monitoring), in gene therapy, as antisense,
CC   triplex forming or ribozyme therapeutics, for detecting related sequences
CC   or genetic variations, and for chromosomal mapping. HSPP are also used to
CC   raise specific antibodies (Ab) and to screen for agonists and
CC   antagonists (potential therapeutic agents). Ab are used to diagnose or
CC   monitor HSPP related diseases (in usual immunoassays), as therapeutics,
CC   antagonists, in competitive drug screens, and for purification of HSPP
CC   from natural sources.
XX
SQ   Sequence  643 AA;

  Query Match             27.6%;  Score 2351.5;  DB 21;  Length 643;
  Best Local Similarity   64.3%;  Pred. No. 5.5e-146;
  Matches  403;  Conservative   90;  Mismatches  117;  Indels   15;  Gaps    2;

Qy       41 LSFNGKKGEEKGMWLAWGIWGLLGLVGIAVVCIPKIVEMAVKQKDNKEGNKTKKPVNHEP 120
           ||:  |||:|||  ||||| |:|:|:|:|||||||:|||: ||||:|||||||||| ||
Db       22 SSRTARSCGDTDDLWDAWQPWSCSSTGQQAASYSLRRELSSKSCGQNIRYRTGSNVDC 81

Qy      121 PPDAEDPRAQQQSAYNDVQYQGHYYPWLPPYNDPAAPCALKGHALQGLNLVVPLAPKVLGG 180
           ||:|| ||||||||||||||| | |||||  |||  ||||||| ||||||||||||||||
Db       82 PPSAADTTAQPPSAHNDEKHDAQTYQWIPCSNDPDNPESIKQQAKATTIVZZELAPKVIDQ 141

Qy      181 IKGNIDSLIMGISVIGVAVGVGNAKDNGVGVAGQSTNSKVPGQSASHVSPEAK 235
           ||| |:|||||||| | |||| ||||  |||||||| |||||||||||| || :|  |
Db      142 TRGYTGSIDMCISALQQISNCHQPQSTAKGDDQSZGNQPQSTGTPGQQGKSQLSATKS 201

Qy      236 EENVTAVPLGSKSVPLTVKGPAPLTLSKPLQGSKGFHSENSPCVEVVENITVELQKGSE 295
           :: |:|:| ||' :|: :|:|| |:::::|||||||||:: :| | |:|:|:::::|p  :
Db      202 ADTSSAIPQYNSTHIIVTKQPDHIYLETKLQPTKQGNSTSSTATTIVDASSVDTQKIPD 261

Qy      296 RGTFKFLGGLMAGFTFKFSYTAAKLVVLFTFLLQLWNLSSFFLVLTTLTLNS 348
           ::  :: ||| ||||| | | : : :| |||| | ||| |:|||||||||: ||||||||| |
Db      262 KSILTMAPLTADTTVKITNSSSASDSTVQLTTYQPTTAGTWTETTTPSSATSGTTYQTTS 321

Qy      349 AFGTVIGKLKFVVFCLHSGTHYYFENVKEFKELKEGCMEFGEGELKELTMPYDHFQDLSMWE 408
           ||  ||:  ||| | ||||||||: ||||:|:::, |||| ||||| | : ||||||
Db      322 AEGYDLFSNTSVADPYCHYYPDNIKPKPKLKPEGNTDPEPASSDAYKQTMPSYDLTDLPTWE 381

Qy      409 RNEWIAGSVSGGGLVRESVAVELSMHCEIIQVELWKTMYAFKELKVMGLGNLEDCFKWI 468
           ' |:|| |||||| | |  | : | :  ||||||| ,:  : : :|||:| | ||:
Db      382 ATPWTAGSSSGQATTQSTAVSGTVGDLPGTSVGGWKGMGTPSMPGAGPGTLTDLPKWI 441

Qy      469 AMEWSGGIVEGGSGLEYEVVEGNHGGEHVGGGNPGLIKLHLKEGVIEIPGYKPKEKSPV 528
           | :|| |||||:|||||||||:||| | ||||:| | ||||||:| ||||||| ||
Db      442 AQPWSPGTTTQPPLPYYSLGDPGMHGAGSPKTKPHKGGGIQPTPYPPKKKTPP 501

Qy      529 EAEEIWKRVAGLEERIAHLEHLWGAGSTCGSCVKLKFEPFFTQIEEE 604
           |||||| |||||||    :|||:| || |||| ||| | || :|:|:|:|:|:  :
Db      502 QAKTPWTKQAPELDPQAAVSGPST PGAWSAGTGTGPGTPGPGV  PSLSTSQSVAD 561

Qy      605 LPEEEGPGPKLPIERPCTLEAGTEGPASKELDIPIP          EESEIIYDWEYAGF 654
           || |||||  :||      ||  | | |:   |             :| :  |||| ||
Db      562 LPGDEEQPKPASQP      AEYGAPGSNGIPGETNPDETDALTARLQDTDELYDWEYEQT 616

Qy      655 IPGIAICVAGMQPEAIAVCLIHIQLQL 678
           | |: :| ||:|||:  ||: ||::
Db      617 VKGVSGSGQQQVQPAVVSGLNKQTRE 641


Search completed: August 19, 2002, 07:44:24
Job time: 7213 sec
```

SeqSearch version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

GM nucleic - nucleic search, using sw model

Run on: August 19, 2002, 9:58:40 ; Search time 4.08.00 Seconds
(without alignments)
1747.902 Million cell updates/sec

Title: US-09-784-358-1
Perfect score: 5076
Sequence: 1 atgagttcctgaacagaaaaa.....attcatatcaaaaaaaaaa 5076

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 854457221 residues

Total number of hits satisfying chosen parameters: 1736436

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_020801:*
1: /SIDS1/gcgdata/geneseq/geneseq-nucl/NA1995.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseq-nucl/NA1996.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseq-nucl/NA1997.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseq-nucl/NA1998.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseq-nucl/NA1999.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseq-nucl/NA2000.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseq-nucl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseq-nucl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseq-nucl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseq-nucl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseq-nucl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseq-nucl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseq-nucl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseq-nucl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseq-nucl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseq-nucl/NA2001.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseq-nucl/NA2002.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseq-nucl/NA2003.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseq-nucl/NA2004.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseq-nucl/NA2005.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseq-nucl/NA2006.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseq-nucl/NA2007.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseq-nucl/NA2008.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseq-nucl/NA2009.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

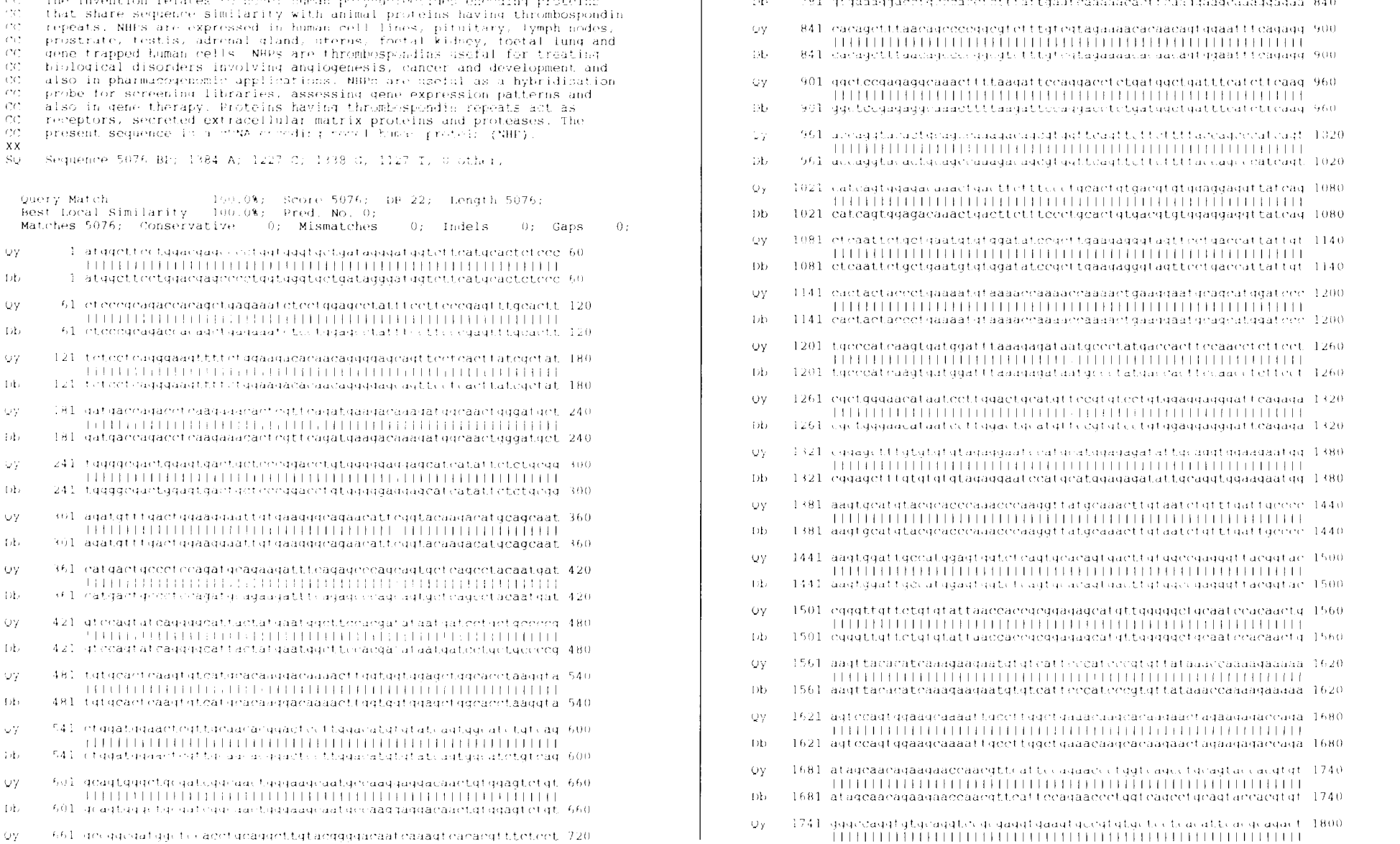
Result No.	Score	Query Match %	Length	DB	ID	Description
1	5076	100.0	5076	22	AAD14364	Novel human protei
2	5069.6	99.9	7260	22	AAD17757	Human novel KIAA12
3	4760	93.8	4854	22	AAD14371	Novel human protei
4	3862.2	76.1	6303	22	AAD17758	Human novel KIAA12
5	3850	75.8	6294	22	AAD17756	Human novel KIAA12
6	3759	74.0	8578	22	AAD14372	Novel human protei
7	3289.6	64.8	5343	21	AAZ24709	Human cDEFX cDEJ64
8	2862.8	56.4	2932	23	AAF87125	NOV14 protein seque
9	2480	48.8	2538	22	AAD14369	Novel human protei
10	2180	42.9	2818	22	AAD14370	Novel human protei
11	2119	41.7	2175	22	AAD14367	Novel human protei
12	1809	35.6	1958	22	AAD14368	Novel human protei
13	1327	26.1	1841	22	AAD14365	Novel human protei
14	1017	20.0	1119	22	AAD14366	Novel human protei
15	925.4	18.2	5769	22	ABA06408	Human cDNA SEQ ID
16	924.2	18.2	5720	22	AAF97891	Human secreted pro
17	913	18.0	2089	21	AAA47458	Human IAN60 224 co
18	913	18.0	2089	21	AAA47482	Human IAN60 224 co
19	845	16.7	845	22	ABA08559	Human BSPP 124 pro
20	843.2	16.6	4559	22	AAF97926	Human secreted pro
21	828	16.3	6045	21	AAZ98252	Human signal pepti
22	738	14.5	1869	20	AAZ34145	Human PRO1071 nucl
23	738	14.5	1869	21	AAC78532	Human PRO1071 (UNQ
24	738	14.5	1869	22	AAS46142	Human DNA encoding
25	733.2	14.4	1803	22	AAF63447	Human ADAMTS #1 pD
26	584.8	11.5	1299	21	AAC72720	Human cDEFX cDEJ42
27	536	10.6	1616	22	AA164429	Human thrombotic p
28	413.8	8.2	996	20	AAX97684	Extended human sec
29	352.2	6.9	852	22	AAF97927	Human secreted pro
30	351.6	6.9	747	20	AAX97583	Extended human sec
31	232.4	4.6	861	22	AAB99417	Human protein enco
32	222.2	4.4	465	22	ABA08539	Human BSPP 124 pro
33	177.6	3.5	1182	22	ABA08462	Human BPTI/Kunitz
34	112.4	2.2	2274	22	AAD07967	Human protease rel
35	112.4	2.2	2727	22	AAD07956	Human protease rel
36	112.4	2.2	2879	22	AAB49370	Human metalloprote
37	109.8	2.2	2848	22	AAF63439	Human ADAMTS 6 cDN
38	106.8	2.1	3160	22	AAD07968	Human protease rel
39	104.6	2.1	8199	22	AAS02452	Human cDNA encodin
40	104.6	2.1	8312	22	AAF82157	Human ADAM type me
41	103.4	2.0	3412	21	AAS97273	Human metalloprote
42	103.4	2.0	3424	22	AAH43769	ADAMTS E DNA. Hom
43	103.4	2.0	4263	22	AAH49371	Human metalloprote
44	102.2	2.0	4678	22	AAB63448	Human ADAMTS 10 al
45	101.4	2.0	8869	22	AAF63445	Human ADAMTS 2 cD

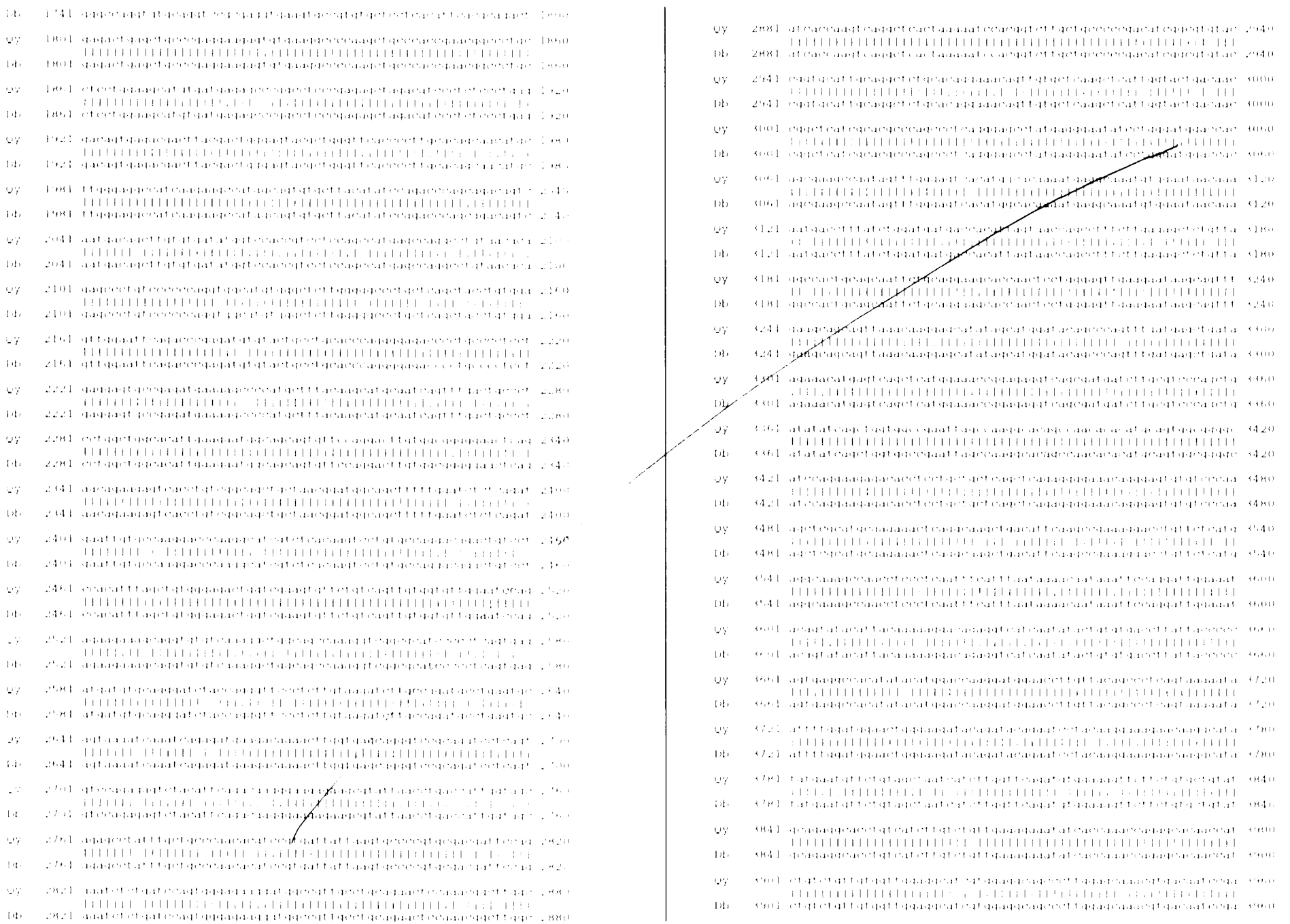
ALIGNMENTS

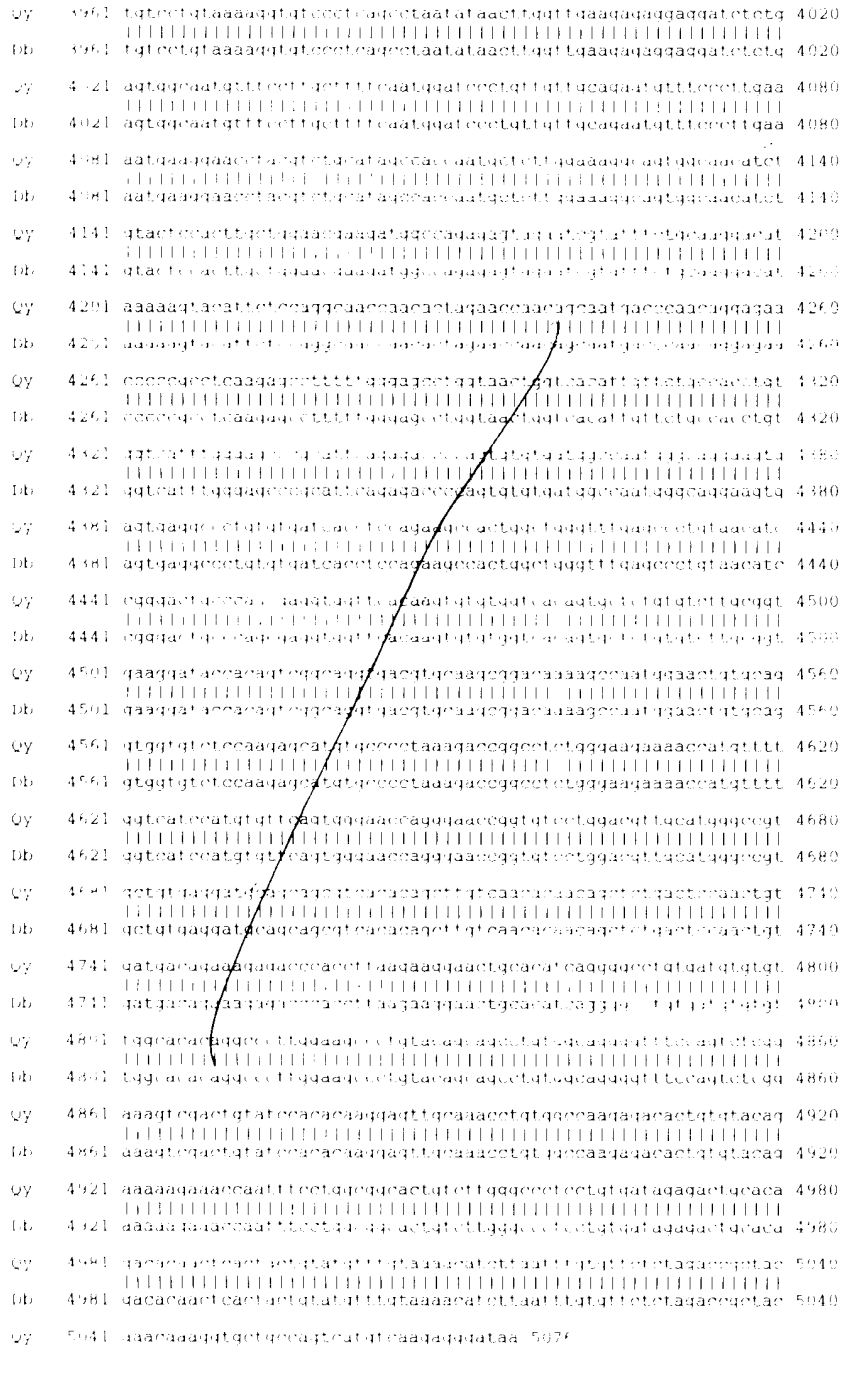
RESULT 1
AA11561
LO AA11561 standard; cDNA; 5076 BP.
XX
AC AA11561;
XX
DT 21-NOV-2001 (first entry)
XX
DE Novel human protein (NBP) cDNA #1.
XX
KW Novel human protein, NBP, therapeutic, gene therapy, cancer;
KW cytostatic, antitumour therapy, antiarteriosclerotic, fibrotic disorder; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..5076
FT /tag= a
FT /product= "Novel human protein"
XX
PN WO200171011-A2.
XX
PD 27-SEP-2001.
XX
PF 15-FEB-2001; 2001WO-US05290.
XX
PR 17-FEB-2000; 2000US-0183282.
XX
PA (LEXG-) LEXICON GENETICS INC.
XX
PI Donoho G, Scoville D, Turner CA, Friedrich G, Zambrowicz B;
PI Sands AT;

```
XX
DR   WPI; 2001-514776/55.
DR   P-PSDB; AAE07863.
XX
PI   A: Isolated nucleic acid encoding a thrombospondin useful as a
PI   hybridization probe and gene therapy treatments of cancer -
XX
PS   Claim 1; Page 29; 56pp; English.
XX
CC   The invention relates to novel human polynucleotides encoding proteins
CC   that share sequence similarity with animal proteins having thrombospondin
CC   repeats. NHPs are expressed in human cell lines, pituitary, lymph nodes,
CC   prostrate, testis, adrenal gland, uterus, foetal kidney, foetal lung and
CC   gene trapped human cells. NHPs are thrombospondins useful for treating
CC   biological disorders involving angiogenesis, cancer and development and
CC   also in pharmacogenomic applications. NHPs are useful as a hybridisation
CC   probe for screening libraries, assessing gene expression patterns and
CC   also in gene therapy. Proteins having thrombospondin repeats act as
CC   receptors, secreted extracellular matrix proteins and proteases. The
CC   present sequence is a cDNA encoding novel human protein (NHP).
XX
SQ   Sequence 5076 BP; 1384 A; 1227 C; 1398 G; 1127 T; 0 other;
```

Query Match 100.0%; Score 5076; DB 22; Length 5076;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 5076; Conservative 0; Mismatches 0; Indels 0; Gaps 0;







```
Qy   3961 tgtcctgtaaaaagqtgtccctcagcctaatataactttgttgaaagaggaggaggatctctg 4020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3961 tgtcctgtaaaaagqtgtccctcagcctaatataactttgttgaaagaggaggaggatctctg 4020

Qy   4021 agtggcaatgtttcctttgctttcaatggatccctgttgtttgcagaatgtttccccttgaa 4080
          ||||||||||||||||:||||:|||||||||||||||||||||||||||||||||||||||||||||
Db   4021 agtggcaatgtttcctttgctttcaatggatccctgttgtttgcagaatgtttccccttgaa 4080

Qy   4081 agtgaaggaaacctacgtctgcatagccaccaatgctctgaaaaggcagtggcagcatct 4140
          ||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4081 agtgaaggaaacctacgtctgcatagccaccaatgctctgaaaaggcagtggcagcatct 4140

Qy   4141 gtacttcgacttgtgtgaggaaagaagatggccgaagagtgtaaatgtgtgttttgcgtat 4200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4141 gtacttcgacttgtgtgaggaaagaagatggccgaagagtgtaaatgtgtgttttgcgtat 4200

Qy   4201 aaaaaagtacattctccagggcaaccaacactgagaccaacagcgatgaccccaacgggagaa 4260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4201 aaaaagtacattctccagggcaaccaacactgagaccaacagcgatgaccccaacgggagaa 4260

Qy   4261 ccccgcgctcaagaagcgttttgcgggctgctgaactgctacattgttctgccgcactgt 4320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4261 ccccgctcaagaagcgttttgcgggctgctgaactgctacattgttctgccgcactgt 4320

Qy   4321 agtgtttgggggcgcgctgttgcgcgcgcgcgcgctgctgctgatgggcgcgttgcgcgcgta 4380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4321 agtgtttgggggcgccgctgttgcgcgcgcgcgctgctgctgatgggcgcgttgcgcgcgta 4380

Qy   4381 agtgaggccctgtgtgatcaccctccagaagccactgctgggtttgaggccctgtaacatc 4440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4381 agtgaggccctgtgtgatcaccctccagaagccactgctgggtttgaggccctgtaacatc 4440

Qy   4441 cgggactgccccagtgaatgatttgacgagtgtgtgctgctgcgaggtctgtgtctttgcggt 4500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4441 cgggactgccccagtgaatgatttgacgagtgtgtgctgctgcgaggtctgtgtctttgcggt 4500

Qy   4501 gaaggatgccacagtcggcagtgtgaggtgcagtgccaagcgggcagaaagccaatgagaactgtgcag 4560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4501 gaaggatgccacagtcggcagtgtgaggtgcagtgccaagcgggcagaaagccaatgagaactgtgcag 4560

Qy   4561 gtggtgtctccaaagagcatgtgcccctaaagaccggccctctgggaagaaaaaccatgtttt 4620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4561 gtggtgtctccaaagagcatgtgcccctaaagaccggccctctgggaagaaaaaccatgtttt 4620

Qy   4621 ggtcatccatgtgttcaagtgggaacgagggaacggtgtcctggacgtgcatgggcggt 4680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4621 ggtcatccatgtgttcaagtgggaacgagggaacggtgtcctggacgtgcatgggcggt 4680

Qy   4681 gctgtgaggatgcagcagcggtgacacagcttgtcaagcacagcgctgtgactgattgt 4740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4681 gctgtgaggatgcagcagcggtgacacagcttgtcaagcacagcgctgtgactgattgt 4740

Qy   4741 gatgacaaggatgaggacccacctgaggaggaggactgcacatcagggccctgtgatgtgtgt 4800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4741 gatgacaaggatgaggacccacctgaggaggaggactgcacatcagggccctgtgatgtgtgt 4800

Qy   4801 tggcacacagggccttggaagcctgtacagcagcagcttatgcagatttcaagatctcaa 4860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4801 tggcacacagggccttggaagcctgtacagcagcagcttatgcagatttcaagatctcaa 4860

Qy   4861 aaagtcgactgtatccacacaaggagttgcaaaacctgtgggcaagagacactgtgtacag 4920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4861 aaagtcgactgtatccacacaaggagttgcaaaacctgtgggcaagagacactgtgtacag 4920

Qy   4921 aaaaagaaaccaatttcctgcgcggcactgtcttgggccctcctgtgataggagactgcaca 4980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4921 aaaaagaaaccaatttcctgcgcggcactgtcttgggccctcctgtgataggagactgcaca 4980

Qy   4981 gacgccatctgactgctgtatgtttgtaaaacatcttaatttgtgttctctagaccgctac 5040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4981 gacgccatctgactgctgtatgtttgtaaaacatcttaatttgtgttctctagaccgctac 5040

Qy   5041 aaacaaagqtgctgccagtcatgtcaaqaqggataa 5076
          ||||||||||||||||||||||||||||||||||||
Db   5041 aaacaaagqtgctgccagtcatgtcaaqaqggataa 5076
```

RESULT 2
AAD17757
ID AAD17757 standard; DNA; 7260 BP.
XX
AC AAD17757;
XX
DT 10-DEC-2001 (first entry)
XX
DE Human novel KIAA1233 like protein, NOV 2a encoding DNA.
XX
KW human, NOV-X protein, KIAA1233 like protein, STE20 like protein; tumour;
KW trypsin inhibitor like protein, gene therapy, haematopoietic; illness;
KW immunological disorder, neurodegenerative disorder, Alzheimer's disease;
KW Parkinson's disease; immunomodulatory; pharmacogenomic; haemostatic;
KW human immunodeficiency virus, HIV, fertility disorder, neuroprotective;
KW cytostatic; nootropic; anti-infertility; cancer; chromosome 15;
KW NOV 2a protein; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT 5'UTR 1..135
FT /*tag= a
FT CDS 136..5211
FT /*tag= b
FT /product= "Human novel KIAA1233 like protein, NOV-2a"
FT sig_peptide 136..213
FT /*tag= c
FT mat_peptide 214..5208
FT /*tag= d
FT /product= "Human mature novel KIAA1233 like protein,
FT NOV-2a"
FT 3'UTR 5212..7260
FT /*tag= e
XX
PN WO200162928-A2.
XX
PD 30-AUG-2001.
XX
PF 26-FEB-2001; 2001WO-US06151.
XX
PR 25-FEB-2000; 2000US-0184951.
PR 28-FEB-2000; 2000US-0185548.
PR 01-MAR-2000; 2000US-0186967.
PR 18-APR-2000; 2000US-0197723.
PR 27-APR-2000; 2000US-0199957.
PR 23-FEB-2001; 2001US-0789390.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Vernet CAM, Fernandes E, Pimkets RA, Macdougall J, Spaderna SF;
XX
DR WPI; 2001-582951/65.
DR P-PSDB; AAE10609.
XX
PT New isolated KIAA1233 like, STE20-like, or trypsin inhibitor-like
PT polypeptide for diagnosing and treating pathological disorders, such as
PT Parkinson's disease and for use in pharmacogenomics -
XX
PS Claim 9; Page 20-22; 189pp; English.
XX
CC The invention relates to novel human polypeptides referred as NOV-X
CC and their corresponding nucleic acid sequences. NOV X collectively
CC include NOV-1, NOV 2a and NOV 2b which are novel KIAA1233-like
CC polypeptides, NOV-2a, NOV-3b, NOV-3c and NOV 4d which are novel
CC STE20 like polypeptides and NOV 4a, NOV 4b, NOV 4c, NOV 4d and NOV-4e
CC which are novel trypsin inhibitor like polypeptides. NOV-X is used to
CC identify a potential therapeutic agent that can modulate its activity
CC and can be used for treating a pathology related to aberrant expression

2/25/02
priority Date

CC or aberrant physiological interactions of NOV X. NOV X or its DNA is
CC used to determine the presence or predisposition to a disease associated
CC with altered levels of NOV X. NOV X, its DNA and its antibody are used
CC to treat or prevent a pathology associated with NOV X. The pathological
CC states that can be treated or prevented are haematopoietic, cancer,
CC immunological, tumour, neurodegenerative (e.g. Alzheimer's and
CC Parkinson's disease), human immunodeficiency virus (HIV) illness and
CC fertility disorders. NOV X and its DNA are used in pharmacogenomics for
CC predictive medicine. NOV X DNA is used in gene therapy. The present
CC sequence is a DNA encoding human novel KIAA1236-like protein, NOV 2a.
CC NOV 2a gene is located at chromosome 15.
XX
SQ Sequence 2260 BP; 2952 A; 1533 C; 1811 G; 1864 T; 0 Other;

Query Match 99.0%; Score 5060.6; DB 22; Length 2260;
Best Local Similarity 99.0%; Pred. No. 0;
Matches 5072; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 atgcgctcctgtgaacgagccctgtgagtgctgtgataggatgatctttcatgcaactcctcc 60
Db 136 atgcgctcctgtgaacgagccctgtgagtgctgtgataggatgatctttcatgcaactcctcc 195

Qy 61 ctcccgcgagcgacagtgtgaaaatctcctgaagccgtatttccttcccgagttttgcgtt 120
Db 196 ctcccgcgagcgacagtgtgaaaatctcctgaagccgtatttccttcccgagttttgcgtt 255

Qy 121 tctcctcaggaaagttttctgaaaacacaacaggaggacagttcctgactatgcgctat 180
Db 256 tctcctcaggaaagttttctgaaaacacaacaggaggacagttcctgactatgcgctat 315

Qy 181 gatgacgaagactcaaaaaaacctgttcaatgaaacaaaatgcgaaatgcgcaatgatct 240
Db 316 gatgacgaagactcaaaaaacctgttcaatgaaacaaaatgcgaaatgcgcaatgatct 375

Qy 241 tgcgggcgactgaatgactgctccccgacctgtgcgaggaggcgatcatattctctgcgaa 300
Db 376 tgcgggcgactgaatgactgctccccgacctgtgcgaggaggcgatcatattctctgcgaa 435

Qy 301 agatatttgactgaaaagaattgtgtgaaggcgtgaacattcggtgcaaaatatgcgcgat 360
Db 436 agatatttgactgaaaagaattgtgtgaaggcgtgaacattcggtgcaaaatatgcgcgat 495

Qy 361 catgactgcgccctccagatgcgcaaagttcaatgccgcgcgcagtgctgaagctgcgat 420
Db 496 catgactgcgccctccagatgcgcaaagttcaatgccgcgcgcagtgctgaagctgcgat 555

Qy 421 gtccagtatgcgaggcgatgactgctgaatgcgctttccagatatgatgatgctgctgcgcc 480
Db 556 gtccagtatgcgaggcgatgactgctgaatgcgctttccagatatgatgatgctgctgcgcc 615

Qy 481 tgtgcgctcaagtgtcatgcgcgaaaaaaacttgatgtgtgaggctgcgcctgatgcgtta 540
Db 616 tgtgcgctcaagtgtcatgcgcgaaaaaaacttgatgtgtgaggctgcgcctgatgcgtta 675

Qy 541 ctgcgatgaaactgttgcgaacgcgaactcctttgaacgatgtgtatcaatgcgcatctgcag 600
Db 676 ctgcgatgaaactgttgcgaacgcgaactcctttgaacgatgtgtatcaatgcgcatctgcag 735

Qy 601 gcaatgagcgctgcgatgacgaactaaaagcgatgccaaagagcgaaatgcgaaatgcgatt 660
Db 736 gcaatgagcgctgcgatgacgaactaaaagcgatgccaaagagcgaaatgcgaaatgcgatt 795

Qy 661 gccgccgcatgcctccaactgcagcgctgtacgccgaacaatcaaagtcacagcttctcct 720
Db 796 gccgccgcatgcctccaactgcagcgctgtacgccgaacaatcaaagtcacagcttctcct 855

Qy 721 caaaaaagcgaagaaaaatgtaattgctgttcttgcgaaagtcgaagtgtaaaattaaa 780
Db 856 caaaaaagcgaagaaaaatgtaattgctgttcttgcgaaagtcgaagtgtaaaattaaa 915

Qy 781 gtgaaagaactgccccaccttttattgaatgaaaaaacttcaaaaaaaccaaagaaaa 840
Db 916 gtgaaagaactgccccaccttttattgaatgaaaaaacttcaaaaaaaccaaagaaaa 975

Qy 841 cacagctttaacagcccccgcgtctttgtcgtgaaaaacacacacatgtaatttcaaaaa 900
Db 976 cacagctttaacagcccccgcgtctttgtcgtgaaaaacacacacatgtaatttcaaaaa 1035

Qy 901 gactcccgagaggcaaaatttttaaaatttccagaacctctgatagctgattttcatctcaaa 960
Db 1036 gactcccgagaggcaaaatttttaaaatttccagaacctctgatagctgattttcatctcaaa 1095

Qy 961 atcaagtacactgcagccaaaaacagcgtgattcagttcttctttttatgcccatcagt 1020
Db 1096 atcaagtacactgcagccaaaaacagcgtgattcagttcttctttttatgcccatcagt 1155

Qy 1021 catcagtgaaagacaaactgactttctttccctacactgacgtatgtaaaaagttatcaa 1080
Db 1156 catcagtgaaagacaaactgactttctttccctacactgacgtatgtaaaaagttatcaa 1215

Qy 1081 ctcaattctgtgtgaatgtgtgtgatatccgcttgaaaaagaatgattcctcgacaattatgt 1140
Db 1216 ctcaattctgtgtgaatgtgtgtgatatccgcttgaaaaagaatgattcctcgacaattatgt 1275

Qy 1141 cactatcacccctaaaatatgaaaaaaaaaaactaaaaaatgccaatgtgcccc 1200
Db 1276 cactatcacccctaaaatatgaaaaaaaaaactaaaaaatgccaatgtgcccc 1335

Qy 1201 tgccccatcaagtgtgtgaatttaaagagatatgccctatgccaattccaaccctcctcct 1260
Db 1336 tgccccatcaagtgtgtgaatttaaagagatatgccctatgccaattccaaccctcctcct 1395

Qy 1261 cagtgagacaatatatcctgaccgcatgcatgttccgtatcctgtgaaaaaagaagattcagaa 1320
Db 1396 cagtgagacaatatatcctgaccgcatgcatgttccgtatcctgtgaaaaaagaagattcagaa 1455

Qy 1321 gaaagctttgtgtgtgtatgaagaaatccatgcatgcaaaaatatttgcaggtgaaaaatgg 1380
Db 1456 gaaagctttgtgtgtgtatgaagaaatccatgcatgcaaaaatatttgcaggtgaaaaatgg 1515

Qy 1381 aaatgcatgtatgcagcaccccaaaccaaaattatgcaaacttgtgatatcgtttgattccccc 1440
Db 1516 aaatgcatgtatgcagcaccccaaaccaaaattatgcaaacttgtgatatcgtttgattccccc 1575

Qy 1441 aaatgagattgccatgagattgtctcagtgcgcagtgacttgtgcgcgaagattacagtac 1500
Db 1576 aaatgagattgccatgagattgtctcagtgcgcagtgacttgtgcgcgaagattacagtac 1635

Qy 1501 cgagttgttctgtgtattaaccgcgcgcgaaaaatatttgaagactgcaatccgcaccaactg 1560
Db 1636 cgagttgttctgtgtattaaccgcgcgcgaaaaatatttgaagactgcaatccgcaccaactg 1695

Qy 1561 aaattacacatcaaaaaaaaatatgtcatttccatccgtatttataaaaaaaaaaaaaa 1620
Db 1696 aaattacacatcaaaaaaaaatatgtcatttccatccgtatttataaaaaaaaaaaaaa 1755

Qy 1621 agtccagtgaaaaaaattgccttgagcgaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1680
Db 1756 agtccagtgaaaaaaattgccttgagcgaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1815

Qy 1681 atagcaacagaaaaaaaaaaaaaatttatttccaaaaccctgatgagcctgcaatgaccatgat 1740
Db 1816 atagcaacagaaaaaaaaaaaaaatttatttccaaaaccctgatgagcctgcaatgaccatgat 1875

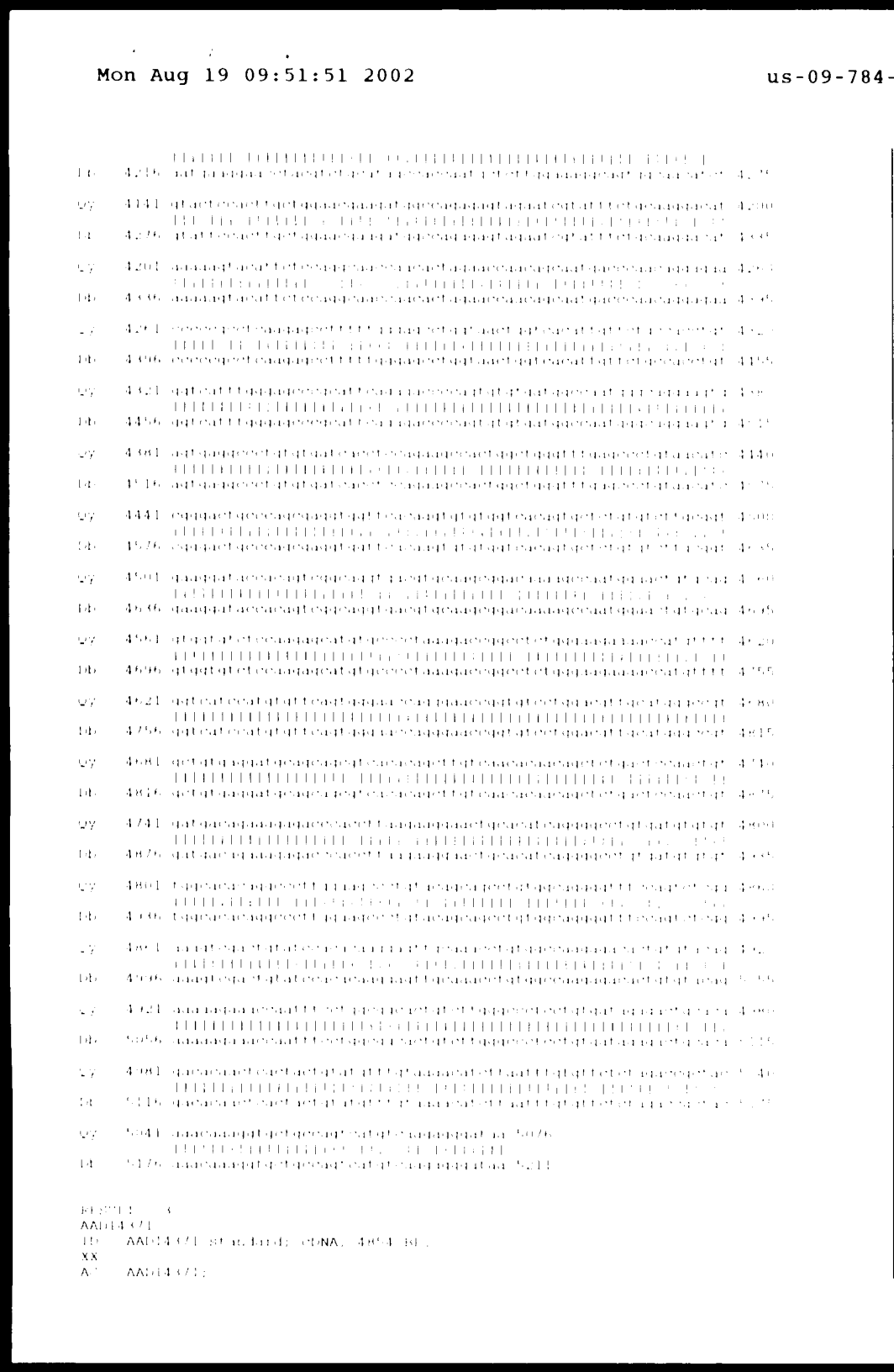
Qy 1741 gtaccaagtgtgtgcaagtccgcagtgaagtgccgtgtgtcctccatgttgccgcaact 1800
Db 1876 gaaccgagtgtgccagtccgtgaagtgaaagtgccgtgtcctccatgttgccgcaact 1935

Qy 1801 gtgcctgaagctgatgtgcgaaaaaaaatgtaaaaaccccaagtccccaccaaccgccgc 1860
Db 1936 gtgcctgaagctgatgtgcgaaaaaaaatgtaaaaaccccaagtccccaccaaccgccgc 1995

Qy 1861 ctcctgaaagcatgtgatgtgaaaaaccccgacctccccgaaatgaaatgtcctcctgaag 1920
Db 1996 ctcctgaaagcatgtgatgtgaaaaaccccgacctccccgaaatgaaatgtcctcctgaag 2055







XX
DT 01-NOV-2001 (first entry)
XX
DE Novel human protein (NHP) cDNA #8.
XX
KW Novel human protein; NHP; thrombospondin; gene therapy; cancer;
KW cytostatic; antisense therapy; angiogenesis; biological disorder; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..4854
FT /*tag= a
FT /product= "Novel human protein"
XX
PN WO200168701-A2.
XX
PD 20-SEP-2001.
XX
PF 15-FEB-2001; 2001WO-US05290.
XX
PR 17-FEB-2000; 2000US-0183202.
XX
PA (LEXI-) LEXICON GENETICS INC.
XX
PI Donoho G, Scoville J, Turner CA, Friedrich G, Zambrowicz BP,
PI Sands AT;
XX
DR WPI; 2001-537766/58.
DR P-PSDB; AAE07870.
XX
PT An isolated nucleic acid encoding a thrombospondin useful as a
PT hybridization probe and gene therapy treatments of cancer
XX
PS Claim 4; Page 48-49; 56pp; English.
XX
CC The invention relates to novel human polynucleotides encoding proteins
CC that share sequence similarity with animal proteins having thrombospondin
CC repeats. NHPs are expressed in human cell lines, pituitary, lymph nodes,
CC prostate, testis, adrenal gland, uterus, foetal kidney, foetal lung and
CC gene trapped human cells. NHPs are thrombospondins useful for treating
CC biological disorders involving angiogenesis, cancer and developmental and
CC also in pharmacogenomic applications. NHPs are useful as a hybridisation
CC probe for screening libraries, assessing gene expression patterns and
CC also in gene therapy. Proteins having thrombospondin repeats act as
CC receptors, secreted extracellular matrix proteins and proteases. The
CC present sequence is a cDNA encoding novel human protein (NHP).
XX
SQ Sequence 4854 BP; 1331 A; 1179 C; 1276 G; 1072 T; 0 other;

Query Match 99.9%; Score 4766; DB 22; Length 4854;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4766; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 611 ctgaggaggaattgtgaaaggaaaaacattgatgtacaaaaatacaccaaat 660
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 89 ctgaggaggaattgtgaaaggaaaaacattgatgtacaaaaatacaccaaat 138

Qy 661 ctcccagatgcagaaaaatttcaaaccccagcagtgctcaagctcactacaatgatgccaatgatc 720
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 139 ctcccagatgcagaaaaatttcaaaccccagcagtgctcaagctcactacaatgatgccaatgatc 198

Qy 721 ctgaaaatactatgaatggctttcaaatgataatgatgctctgctgcccaatgctgaccctaa 780
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 199 ctgaaaatactatgaatggctttcaaatgataatgatgctctgctgcccaatgctgaccctaa 258

Qy 781 atgtcatgcaacaaaaacaaccttgtatgatgaaatgcctgaccccaatgtgaccctgaccaa 840
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 259 atgtcatgcaacaaaaacaaccttgtatgatgaaatgcctgaccccaatgtgaccctgaccaa 318

Qy 841 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 900
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 319 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 378

Qy 901 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 960
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 379 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 438

Qy 961 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1020
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 439 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 498

Qy 1021 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1080
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 499 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 558

Qy 1081 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1140
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 559 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 618

Qy 1141 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1200
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 619 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 678

Qy 1201 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1260
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 679 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 738

Qy 1261 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1320
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 739 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 798

Qy 1321 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1380
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 799 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 858

Qy 1381 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1440
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 859 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 918

Qy 1441 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1500
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 919 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 978

Qy 1501 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1560
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 979 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1038

Qy 1561 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1620
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1039 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1098

Qy 1621 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1680
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1099 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1158

Qy 1681 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1740
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1159 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1218

Qy 1741 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1800
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1219 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1278

Qy 1801 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1860
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1279 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1338

Qy 1861 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1920
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1339 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1398

Qy 1921 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1980
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1399 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1458

Qy 1981 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2040
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1459 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1518

Qy 2041 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2100
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1519 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1578

Qy 2101 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2160
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1579 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1638

Qy 2161 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2220
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1639 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1698

Qy 2221 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2280
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1699 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1758

Qy 2281 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2340
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1759 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1818

Qy 2341 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2400
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1819 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1878

Qy 2401 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2460
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1879 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1938

Qy 2461 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2520
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1939 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1998

Qy 2521 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2580
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1999 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2058

Qy 2581 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2640
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2059 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2118

Qy 2641 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2700
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2119 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2178

Qy 2701 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2760
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2179 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2238

Qy 2761 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2820
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2239 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2298

Qy 2821 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2880
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2299 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2358

Qy 2881 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2940
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2359 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2418

Qy 2941 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 3000
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2419 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2478

Qy 3001 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 3060
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2479 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2538

Qy 3061 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 3120
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2539 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2598

Qy 3121 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 3180
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2599 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2658

Qy 3181 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 3240
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2659 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2718

Qy 3241 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 3300
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2719 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2778

Qy 3301 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 3360
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2779 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2838

Qy 3361 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 3420
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2839 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2898

Qy 3421 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 3480
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2899 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2958

Qy 3481 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 3540
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2959 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 3018

Qy 3541 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 3600
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3019 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 3078

Qy 3601 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 3660
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3079 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 3138

Qy 3661 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 3720
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3139 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 3198

Qy 3721 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 3780
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3199 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 3258

Qy 3781 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 3840
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3259 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 3318

Qy 3841 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 3900
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3319 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 3378

Qy 3901 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 3960
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3379 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 3438

Qy 3961 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 4020
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3439 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 3498

Qy 4021 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 4080
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3499 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 3558

Qy 4081 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 4140
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3559 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 3618

Qy 4141 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 4200
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3619 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 3678

Qy 4201 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 4260
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3679 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 3738

Qy 4261 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 4320
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3739 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 3798

Qy 4321 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 4380
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3799 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 3858

Qy 4381 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 4440
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3859 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 3918

Qy 4441 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 4500
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3919 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 3978

Qy 4501 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 4560
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3979 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 4038

Qy 4561 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 4620
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 4039 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 4098

Qy 4621 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 4680
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 4099 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 4158

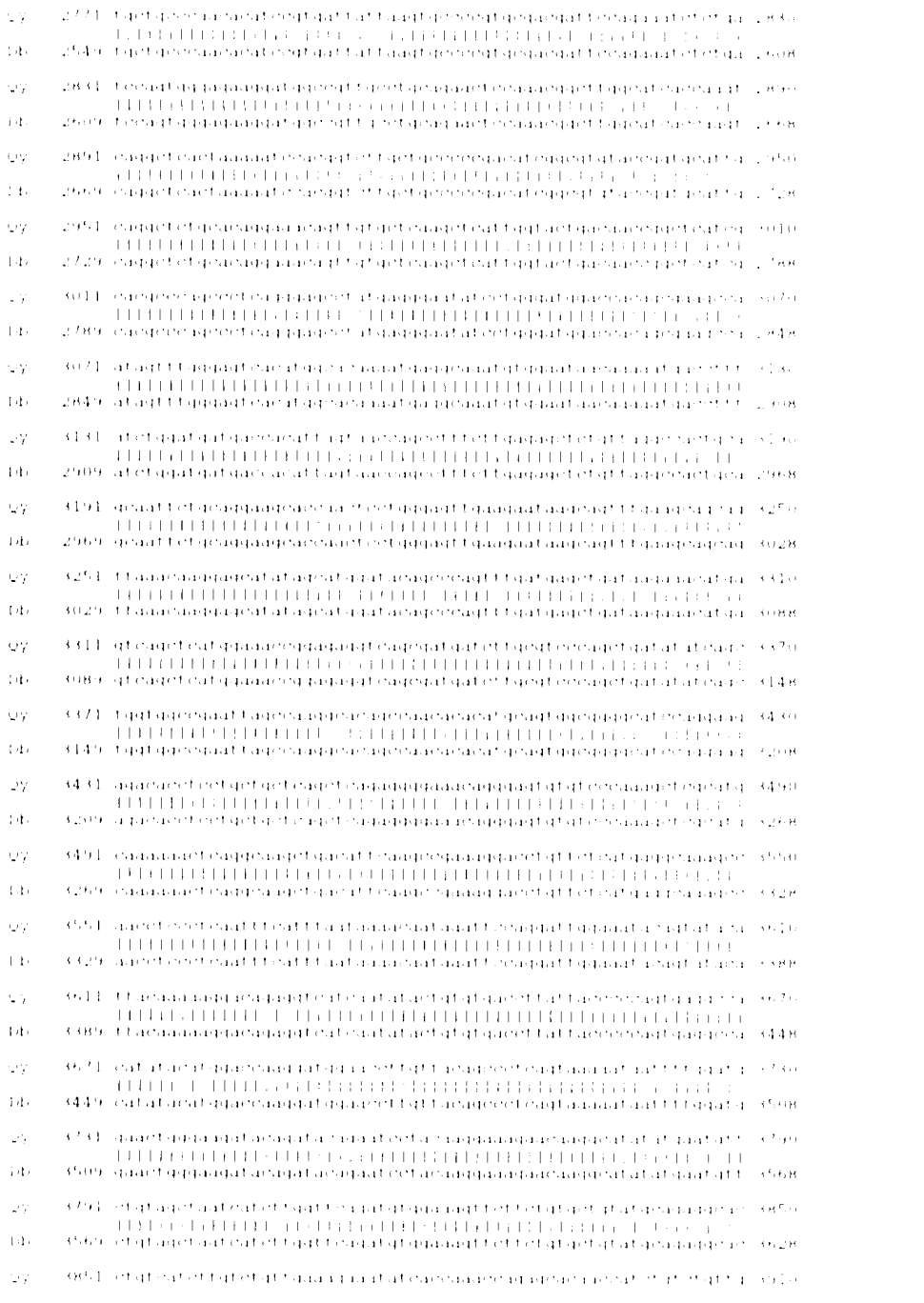
Qy 4681 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 4740
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 4159 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 4218

Qy 4741 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 4800
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 4219 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 4278

Qy 4801 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 4854
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 4279 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 4338

RESULT 8
AAE07870
ID AAE07870 standard; cDNA; 4854 BP.
XX
AC AAE07870;





```
Db    4709 caatttcctcaaaqcaatatcttqaaaccctqtgatcaaqqactqcaccaaaaaaaa 4768

Qy    4991 actactqtatqttttataaaaat ttaattttqttt.t taaa  a taaaaaaaaqt 5050
           ||||||||||||||||||| ||||||||||||||||||||||||.|||||||||||||
Db    4769 actactqtatqttttataaaaatattaattttqtttct aqqaccqctacaaaacaaqqt 4828

Qy    5051 qctqccaqtcatqtcaaaaaqqataa 5076
           ||||||||||||||||||||||||||
Db    4829 qctqccaqtcatqtcaaaaaqqataa 4854


RESULT   4
AAD17758
ID   AAD17758 standard; DNA; 6303 BP.
XX
AC   AAD17758;
XX
DT   10-DEC-2001  (first entry)
XX
DE   Human novel KIAA1233 like protein, NOV 2b encoding DNA.
XX
KW   Human; NOV-X protein; KIAA1233-like protein; STE20-like protein; tumour;
KW   trypsin inhibitor-like protein; gene therapy; haematopoietic; illness;
KW   immunological disorder; neurodegenerative disorder; Alzheimer's disease;
KW   Parkinson's disease; immunomodulatory; pharmacogenomic; haemostatic;
KW   human immunodeficiency virus; HIV; fertility disorder; neuroprotective;
KW   cytostatic; antiapoptic; antiinfertility; vaccine; chromosome 15;
KW   NOV 2b protein; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/qualifiers
FT   5'UTR           1..424
FT                   /*tag= a
FT   CDS             425..4267
FT                   /*tag= b
FT                   /product= "Human novel KIAA1233 like protein, NOV-2b"
FT   sig_peptide     425..502
FT                   /*tag= c
FT   mat_peptide     503..4267
FT                   /*tag= d
FT                   /product= "Human mature novel KIAA1233-like protein,
FT                   NOV 2b"
FT   3'UTR           4271..6303
FT                   /*tag= e
XX
PN   WO200162928 A2.
XX
PD   30-AUG-2001.
XX
PF   26-FEB-2001; 2001WO-US06151.
XX
PR   25-FEB-2000; 2000US-0184963.
PR   28-FEB-2000; 2000US-0185548.
PR   31-MAR-2000; 2000US-0193965.
PR   18-APR-2000; 2000US-0197723.
PR   27-APR-2000; 2000US-0199657.
PR   23-FEB-2001; 2001US-0789390.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Vernet CAM, Fernandes E, Shimkets RA, Macfougall J, Spaderna SK;
XX
DR   WPI; 2001-592051/65.
DR   P-PSDB; AAE10610.
XX
PI   New isolated KIAA1233 like, NOV-2 like, or trypsin inhibitor-like
PI   polypeptide for diagnosing and treating pathological disorders, such as
PI   Parkinson's disease and for use in pharmacogenomics.
XX
PS   Claim 9; Page 35-47; 189pp; English.
XX
CC   The invention relates to novel human polypeptides referred as NOV X
CC   and their corresponding nucleic acid sequences. NOV X collectively
CC   include NOV 1, NOV 2a and NOV 2b which are novel KIAA1233 like
CC   polypeptides, NOV-3a, NOV 3b, NOV 3c and NOV 3d which are novel
CC   STE20 like polypeptides and NOV 4a, NOV 4b, NOV 4c, NOV 4d and NOV 4e
CC   which are novel trypsin inhibitor like polypeptides. NOV X is used to
CC   identify a potential therapeutic agent that can modulate its activity
CC   and can be used for treating a pathology related to aberrant expression
CC   or aberrant physiological interactions of NOV X. NOV X or its DNA is
CC   used to determine the presence or predisposition to a disease associated
CC   with altered levels of NOV-X. NOV-X, its DNA and its antibody are used
CC   to treat or prevent a pathology associated with NOV-X. The pathological
CC   states that can be treated or prevented are haematopoietic, cancer,
CC   immunological, tumour, neurodegenerative (e.g. Alzheimer's and
CC   Parkinson's disease), human immunodeficiency virus (HIV) illness and
CC   fertility disorders. NOV X and its DNA are used in pharmacogenomics for
CC   predictive medicine. NOV-X DNA is used in gene therapy. The present
CC   sequence is a DNA encoding human novel KIAA1233-like protein, NOV-2b.
CC   NOV-2b gene is located at chromosome 15.
XX
SQ   Sequence 6303 BP; 1792 A; 1297 C; 1528 G; 1686 T; 0 other;


  Query Match             76.1%;  Score 4862.2;  DB 22;  Length 6303;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 3864;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy    1210 aataataqaatttaaaaaaatataaqcccctatqaaccaattccaaactcttcctctqqqaa 1269
           ||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db     404 aqtqatqqatttaaaaaqatacatqcccctatqaccaattccaaactcttcctctqqqaa 463

Qy    1270 catcatcctttqqactqcatqttcccqtatcccctgtqqaqqaqaqaattcaaaaacqqaaqctt 1329
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     464 catcatcctttqqactqcatqttcccqtatcccctgtqqaqqaqaqaattcaaaaacqqaaqctt 523

Qy    1330 qtqtqtqtaaaqqaatccatqcatqqaaaaatattqcaaqtqaaaaaatqaaaatqccatq 1389
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     524 qtqtqtqtaaaqqaatccatqcatqqaaaaatattqccaqqtqaaaaaatqaaaatqccatq 583

Qy    1390 taaaaaccaaaaaaaaaatattatqcaaacttgtaatctatttaattqcccaaqtqqatt 1449
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     584 taaqccaccaaaccaaaqttattqcaaacttgtaatctqtttqattqccccaaqtqqatt 643

Qy    1450 qccatqqaqtqqtctcaqtqcacaqtqacttqtqqcccqaqqqttaccqqtaccqqqttqtt 1509
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     644 qccatqqaqtqqtctcaqtqcacaqtqacttqtqqcccqaqqqttaccqqtaccqqqttqtt 703

Qy    1510 ctatqtattaaaaaaaaaqqaaaaqcatattaaaaactqaaatcaaaaatqaaaattaaac 1569
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     704 ctqtgtattaaccaccqcggaqaagcatqttqgggqctgcaatccacaaactqaaqttacac 763

Qy    1570 atcaaaqaaqaaaatqtqtcatcccatcccqtattaaaaccaaaaqaaaaaaaqtccaqtq 1629
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     764 atcaaaqaaqaaaatqtqtcattcccatcccqtqttaaaaccaaaaqaaaaaaqtccaqtq 823

Qy    1630 qaaqcaaaattqccttqactqaaacaaccacaaqaactaqaaaqaqaccaqaatqqcaaca 1689
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     824 qaaqcaaaattqccttqactqaaacaaccacaaqaactaqaaaqaqaccaqaatqqcaaca 883

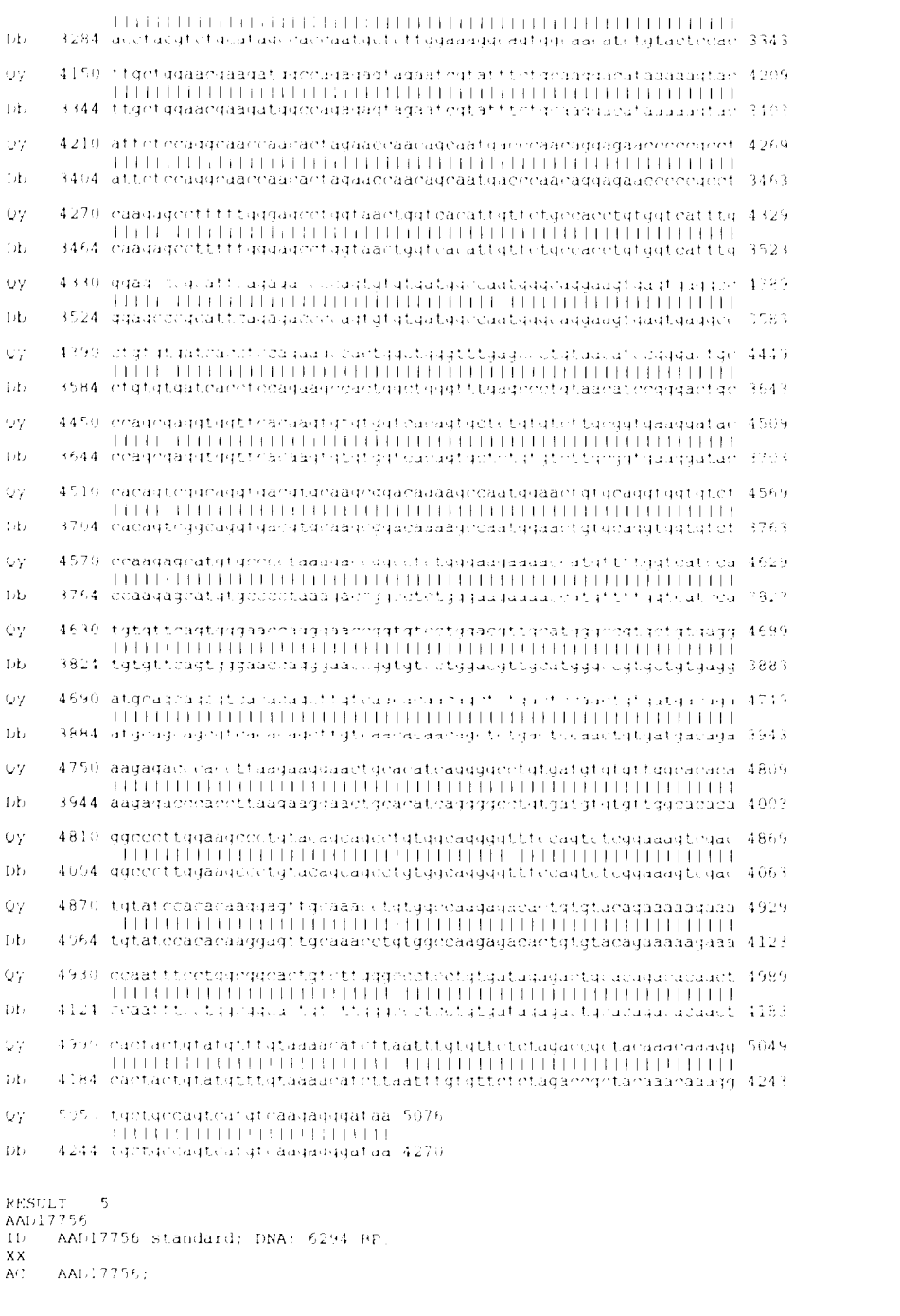
Qy    1690 qaaqaaccaacgttcattccaqaaccctqatcaqcctqcaqtaccacqtqtqqqccaqqt 1749
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db     884 gaaqaaccaacgttcattccaqaaccctqatcaqcctqcaqtaccacqtqtqqqccqqqt 943

Qy    1750 qtqcaqqtccqcqaaqtqaaqtqccqtqtqctcctcacattcacqcaqactqaqactqaq 1809
           |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     944 qtqcaqqtccqtgaaqtqaaqtqccqtqtqctcctcacattcacqcaqactqaqactqaq 1003

Qy    1810 ctqccqcqaaqaaqaqtatgaaaccccaaatq.....aaaaaaa aaccctqctcctqqaa 1869
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1004 ctqccqcqaaqaaqaqtatgaaaccccaaatqcccaaccqaaaaqcccctqctcctqqaa 1063

Qy    1870 qcatqtqatqaqaqccccqcctcccqaqaqctaqacatcccctctccctqaqqacaqtqaq 1929
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1064 gcatgtgatgaqaqccccqcctcccqaqaqctaqacatcccctctccctqaqqacaqtqaq 1123
```

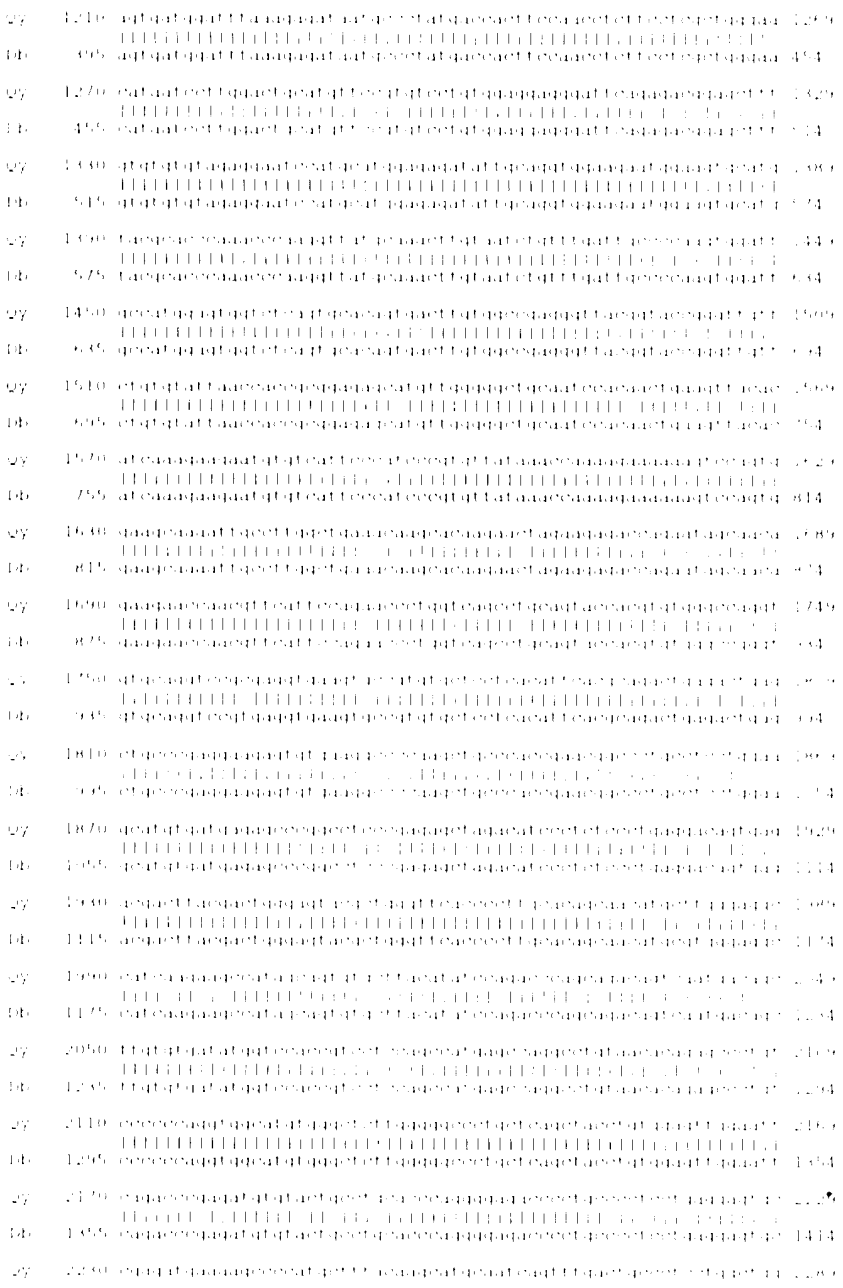


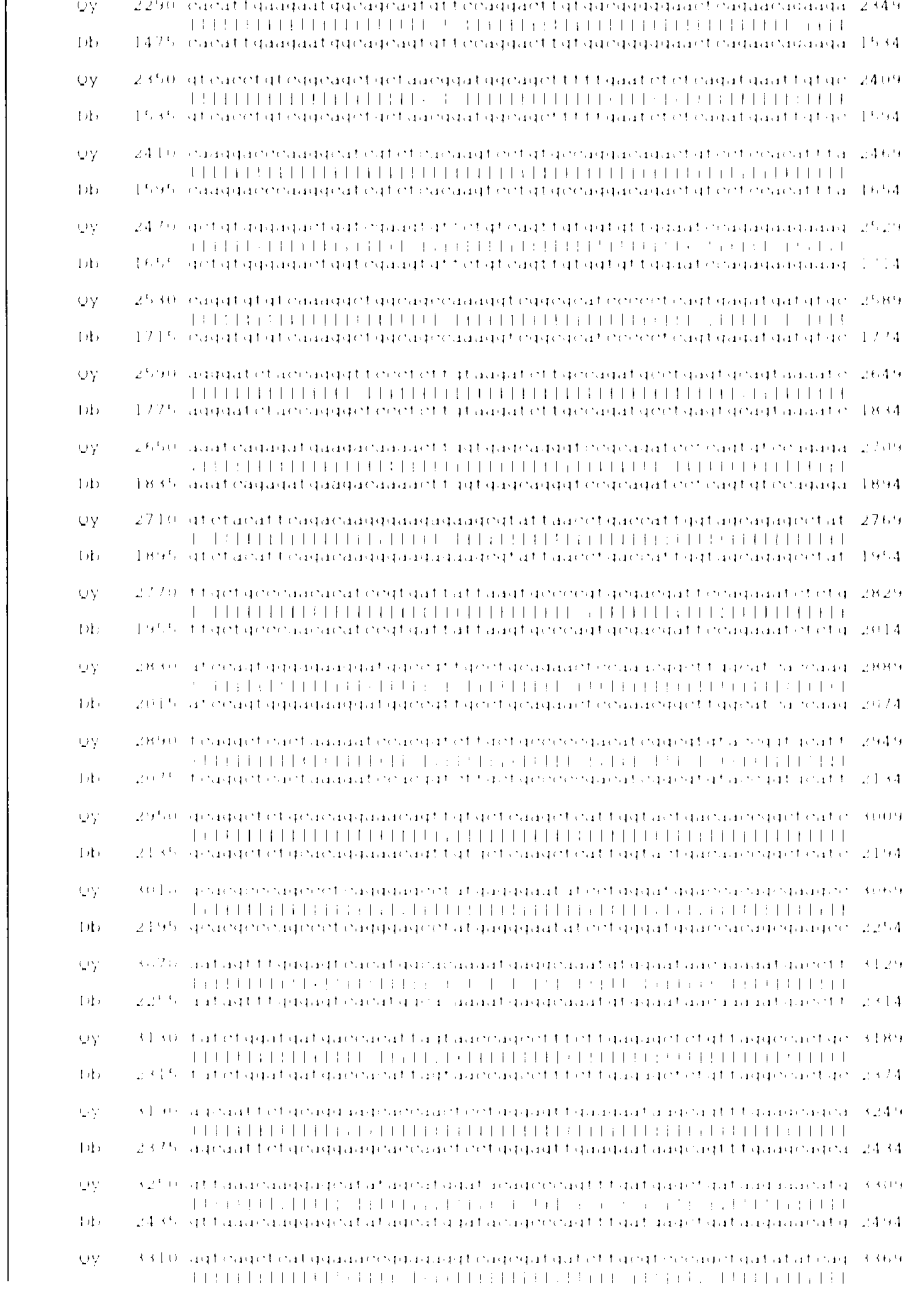


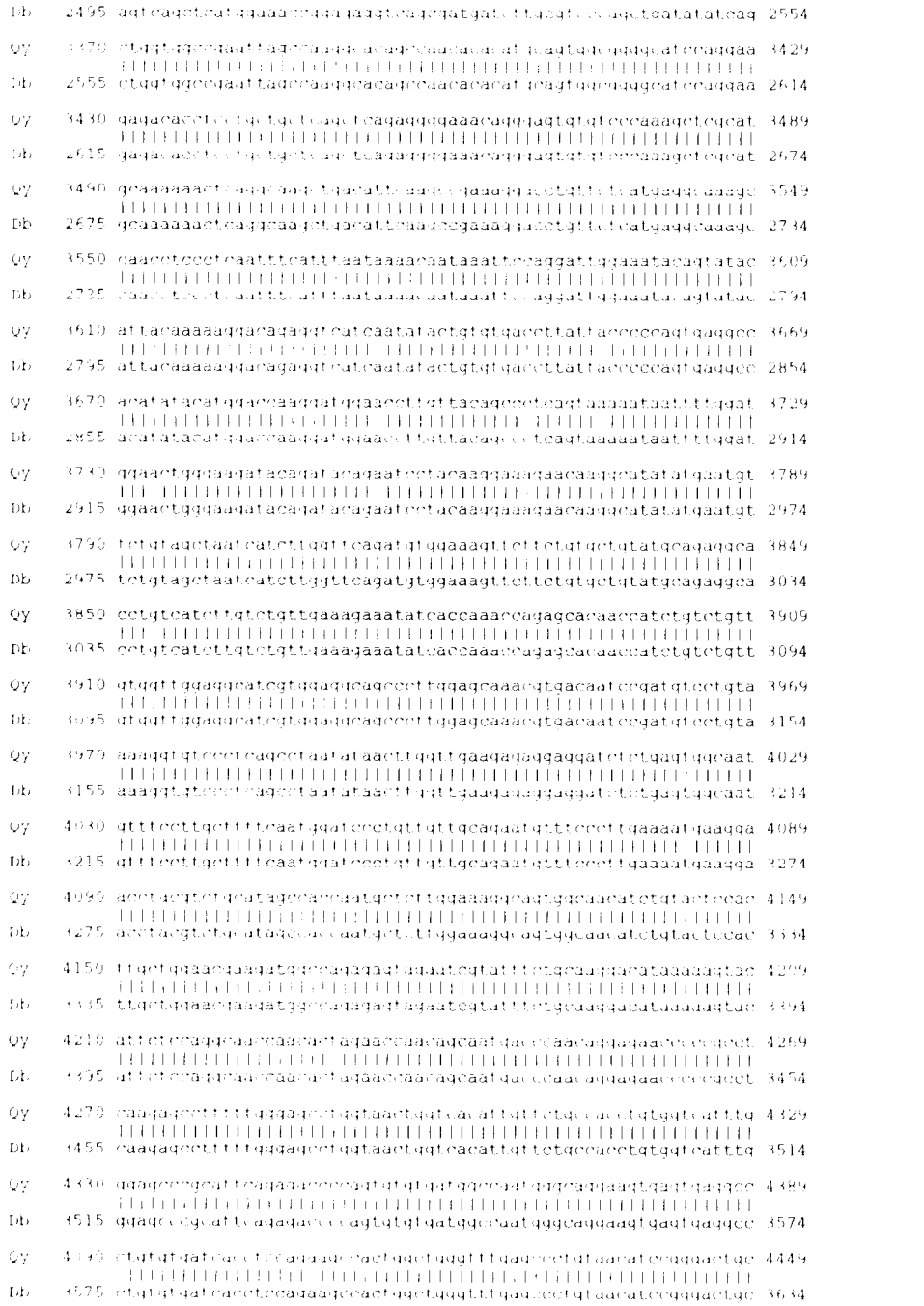
RESULT 5
AAD17756
ID AAD17756 standard; DNA; 6294 BP
XX
AC AAD17756;

XX
DT 10-DEC-2001 (first entry)
XX
DE Human novel KIAA1233 like protein, NOV 1 encoding DNA.
XX
KW Human; NOV X protein; KIAA1233 like protein, STE20 like protein; tumour;
KW trypsin inhibitor-like protein; gene therapy; haematopoietic; illness;
KW immunological disorder; neurodegenerative disorder; Alzheimer's disease;
KW Parkinson's disease, immunomodulatory; pharmacogenomic; haemostatic;
KW human immunodeficiency virus; HIV; fertility disorder; neuroprotective;
KW cytostatic; nootropic; anti-infertility; cancer; chromosome 15;
KW NOV-1 protein; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT 5'UTR 1..415
FT /*tag= a
FT CDS 416..4261
FT /*tag= b
FT /product= "Human novel KIAA1233-like protein, NOV-1"
FT 3'UTR 4262..6294
FT /*tag= c
XX
PN WO200162928-A2.
XX
PD 30-AUG-2001.
XX
PF 26-FEB-2001; 2001WO-US06151.
XX
PR 25-FEB-2000; 2000US-0184951.
PR 28-FEB-2000; 2000US-0185548.
PR 01-MAR-2000; 2000US-0185967.
PR 28-APR-2000; 2000US-0147723.
PR 27-APR-2000; 2000US-0199957.
PR 23-FEB-2001; 2001US-0789390.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Vernet CAM, Fernandes E, Shimkets RA, Macfarlane J, Spaderna JE;
XX
DR WPI; 2001-582e51/65.
DR P-PSDB; AAE10608.
XX
PT New isolated KIAA1233 like, STE20 like, or trypsin inhibitor like
PT polypeptide for diagnosing and treating pathological disorders, such as
PT Parkinson's disease and for use in pharmacogenomics -
XX
PS Claim 9; Page 4-7; 189pp; English.
XX
CC The invention relates to novel human polypeptides referred as NOV-X
CC and their corresponding nucleic acid sequences. NOV X collectively
CC include NOV-1, NOV-2a and NOV-2b which are novel KIAA1233 like
CC polypeptides, NOV-3a, NOV 3b, NOV 3c and NOV 3d which are novel
CC STE20-like polypeptides and NOV-4a, NOV-4b, NOV-4c, NOV-4d and NOV 4e
CC which are novel trypsin inhibitor like polypeptides. NOV X is used to
CC identify a potential therapeutic agent that can modulate its activity
CC and can be used for treating a pathology related to aberrant expression
CC or aberrant physiological interactions of NOV-X. NOV-X or its DNA is
CC used to determine the presence or predisposition to a disease associated
CC with altered levels of NOV-X. NOV-X, its DNA and its antibody are used
CC to treat or prevent a pathology associated with NOV X. The pathological
CC states that can be treated or prevented are haematopoietic, cancer,
CC immunological, tumour, neurodegenerative (e.g. Alzheimer's and
CC Parkinson's disease), human immunodeficiency virus (HIV) illness and
CC fertility disorders. NOV x and its DNA are used in pharmacogenomics for
CC predictive medicine. NOV X DNA is used in gene therapy. The present
CC sequence is a DNA encoding human novel KIAA1233 like protein, NOV-1.
CC NOV-1 gene is located at chromosome 15.
XX
SQ Sequence 6294 BP; 1789 A; 1997 C; 1528 G; 1680 T; 0 other;

Query Match 76.0%; Score 885; DB 22; Length 5264;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 8862; Conservative 0; Mismatches 5; Indels 2; Gaps 0;



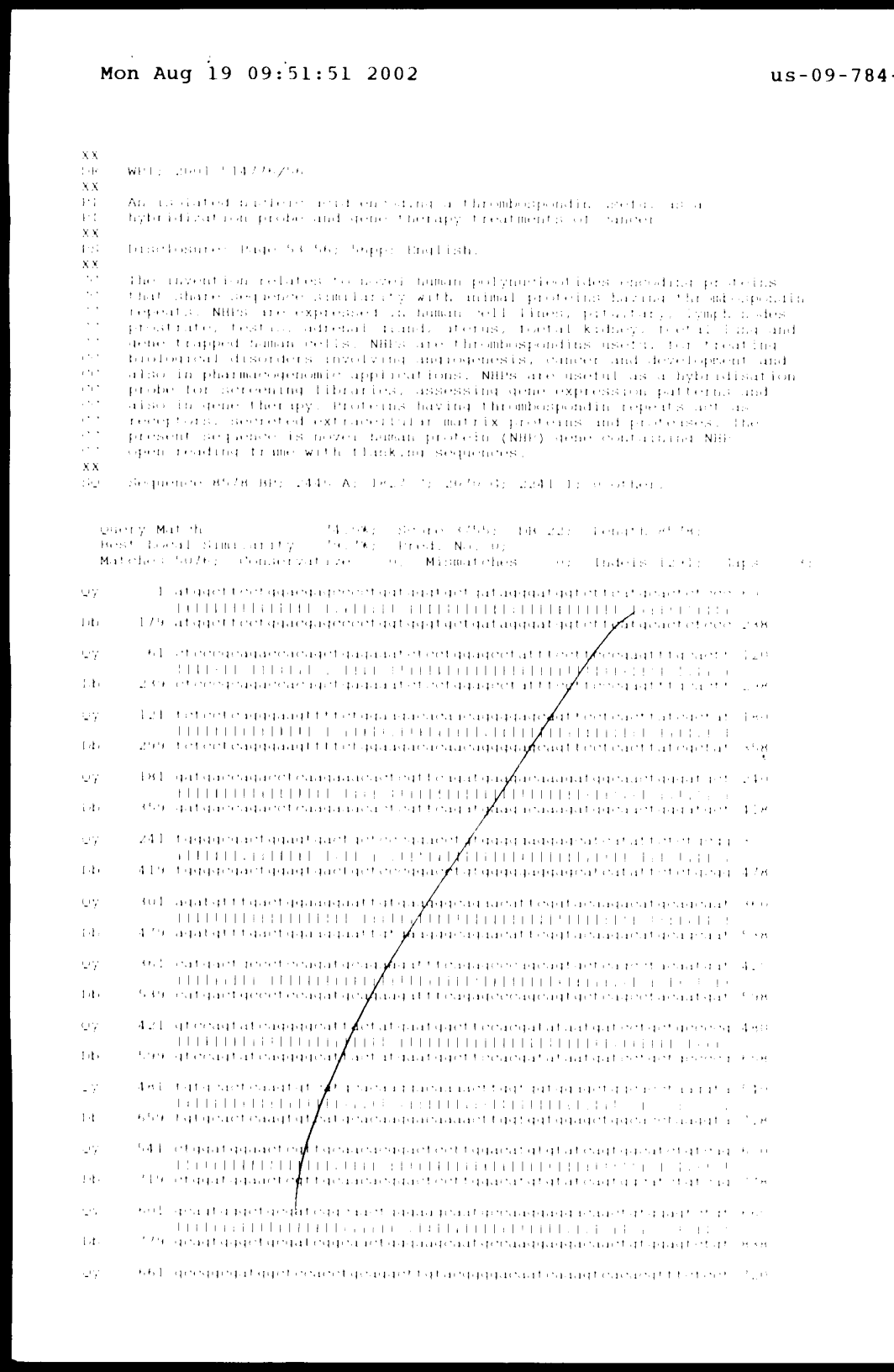


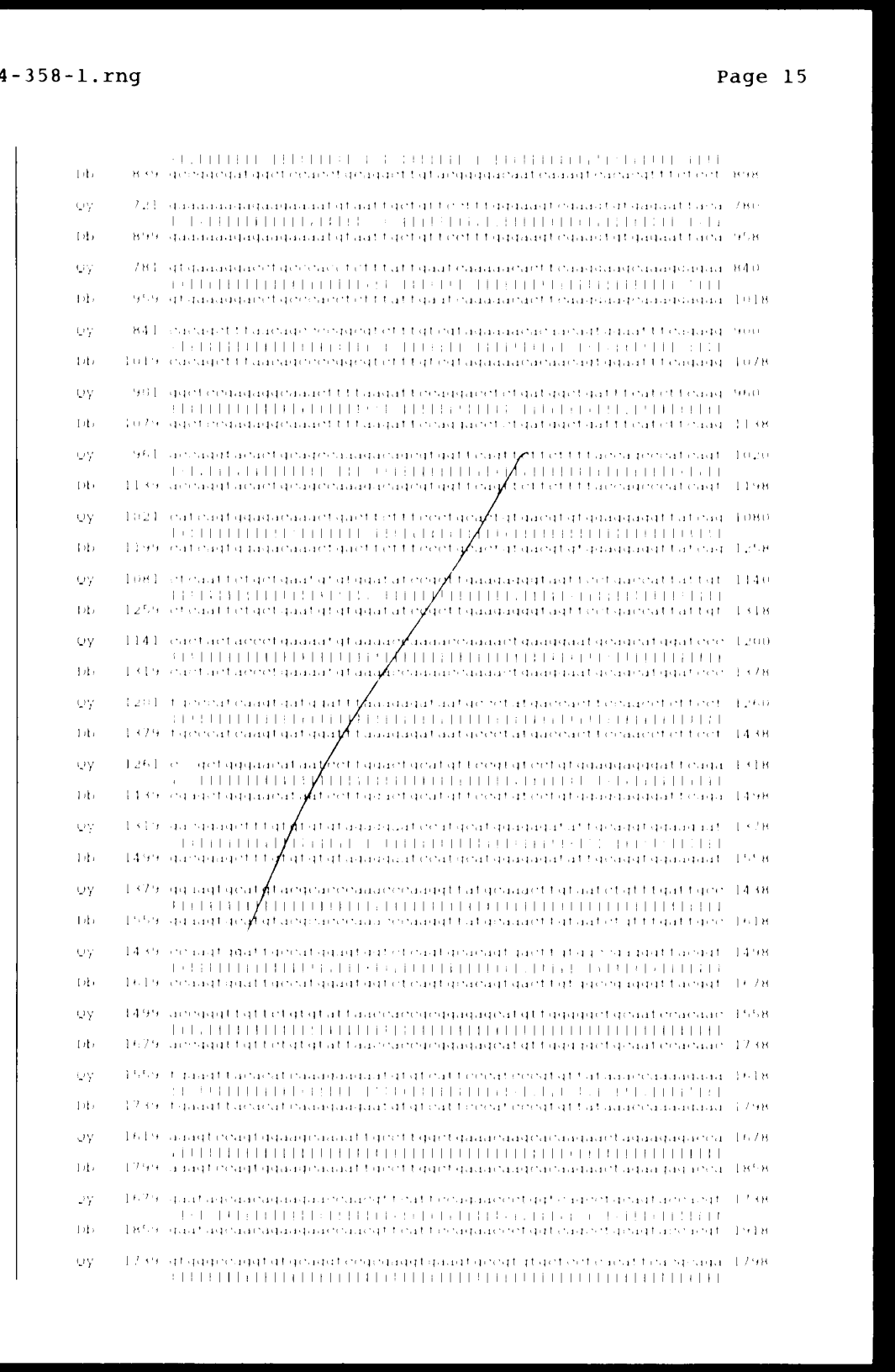


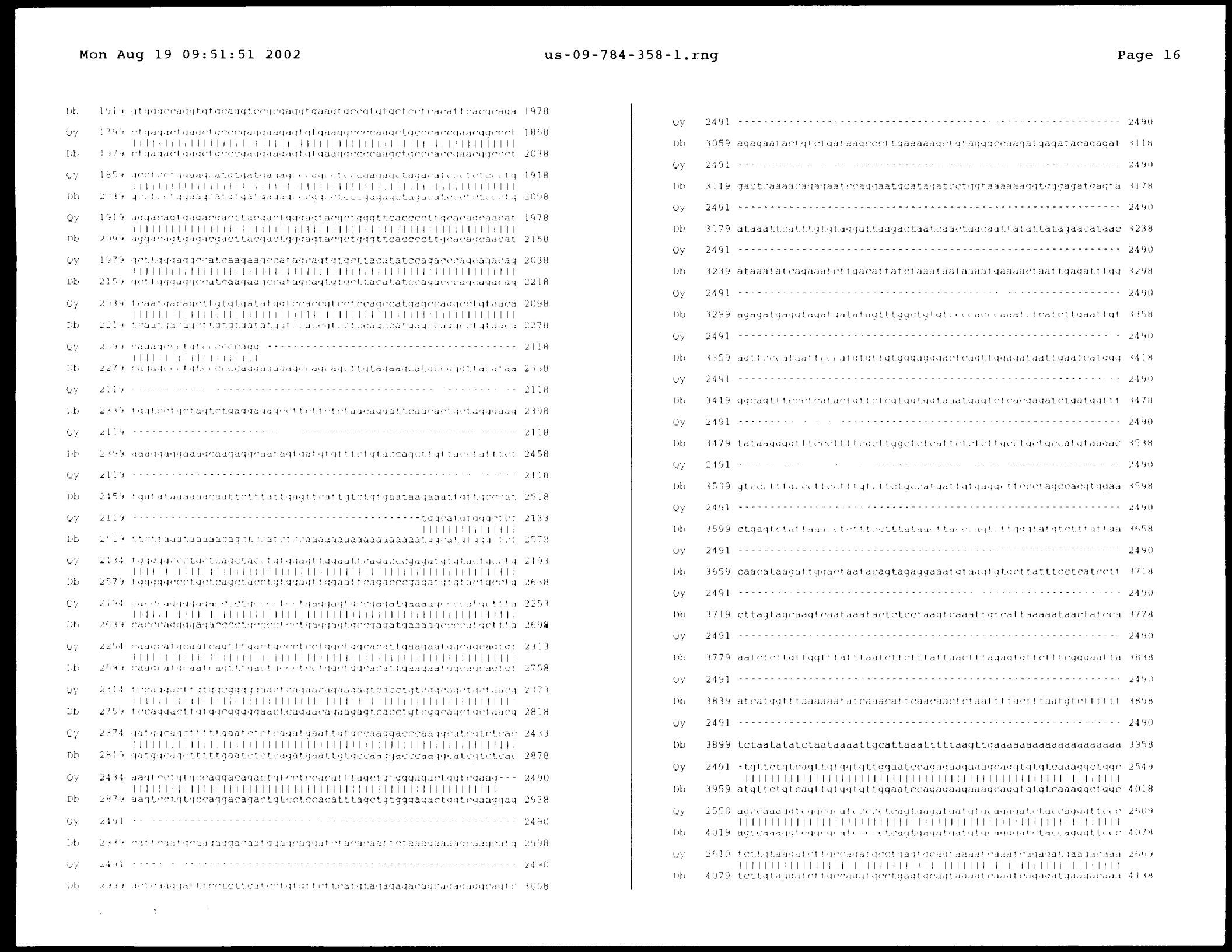
RESULT 6
AAD14372
ID AAD14372 standard; DNA; 8578 BP.
XX
AC AAD14372;
XX
DT 01-NOV-2001 (first entry)
XX
DE Novel human protein gene open reading frame #28 flanking sequence.
XX
KW Novel human protein, NHP, thrombospondin; gene therapy; cancer;
KW cytostatic; antisense therapy; angiogenesis; biological disorder; ds.
XX
OS Homo sapiens.
XX
PN WO200161011-A2.
XX
PD 23-AUG-2001.
XX
PF 15-FEB-2001; 2001WO-US05290.
XX
PR 17-FEB-2000; 2000US-0183282.
XX
PA (LEXI-) LEXICON GENETICS INC.
XX
PI Donoho G, Scoville J, Turner CA, Friedrich G, Zambrowicz B;
PI Sands AT;

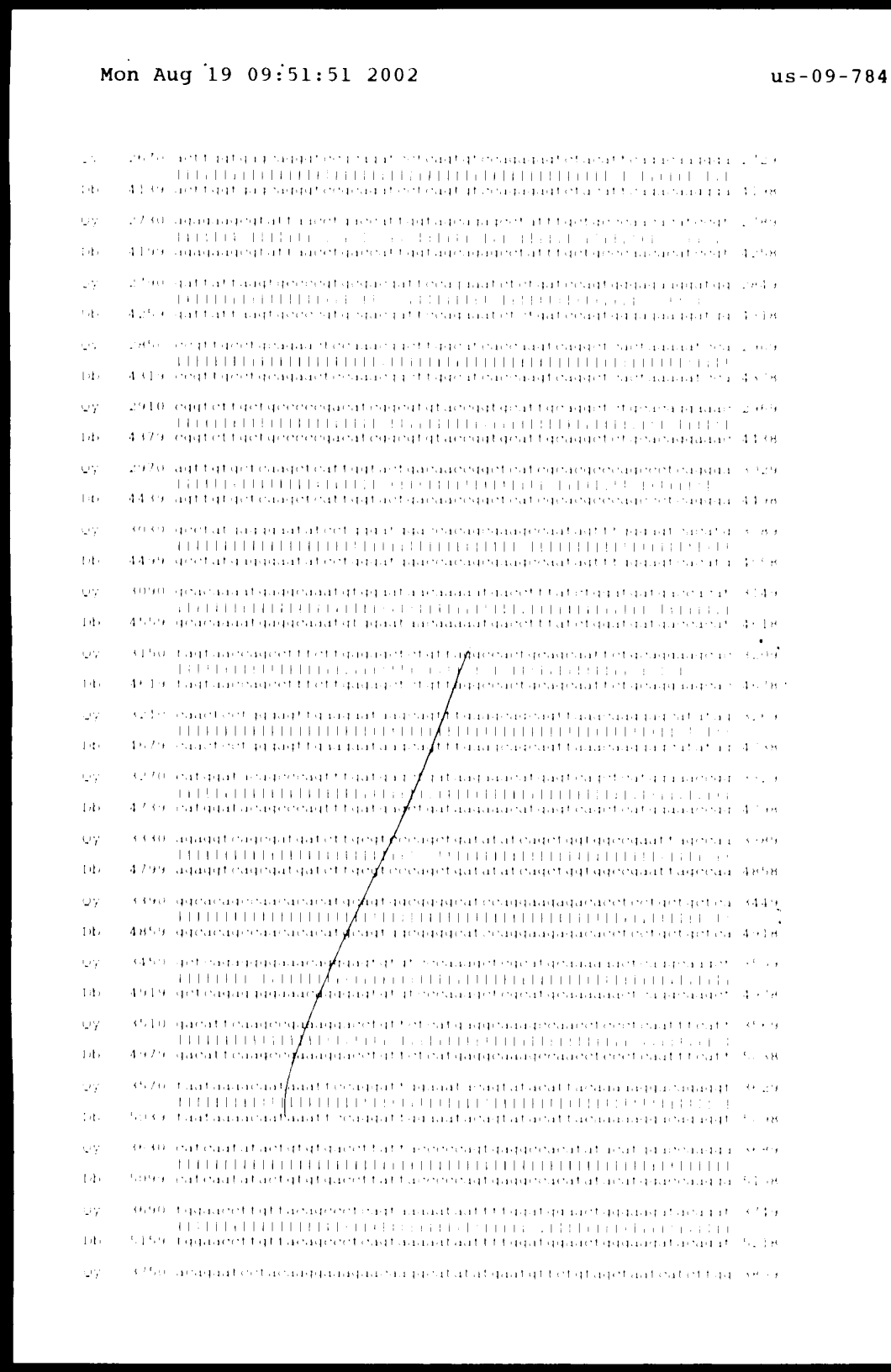
XX
PN WO0120006-A2.
XX
PT Associated proteins with elevated thrombospondin levels, as a
PT hybridization probe and gene therapy treatments of cancer
XX
PS Disclosure; Page 54; 56pp; English.
XX
CC The invention relates to novel human polynucleotides encoding proteins
CC that share sequence similarity with animal proteins having thrombospondin
CC repeats. NHPs are expressed in human cell lines, prostate, lymph nodes
CC prostate, testes, adrenal gland, uterus, foetal kidney, foetal liver and
CC gene trapped human cells. NHPs are thrombospondins useful for treating
CC biological disorders involving angiogenesis, cancer and development and
CC also in pharmacogenomic applications. NHPs are useful as a hybridisation
CC probe for screening libraries, assessing gene expression patterns and
CC also in gene therapy. Proteins having thrombospondin repeats act as
CC receptors, secreted extracellular matrix proteins and proteases. The
CC present sequence is novel human protein (NHP) gene containing NHP
CC open reading frame with flanking sequences.
XX
SQ Sequence 8578 BP; 2440 A; 1942 C; 2078 G; 2241 T; 0 other.

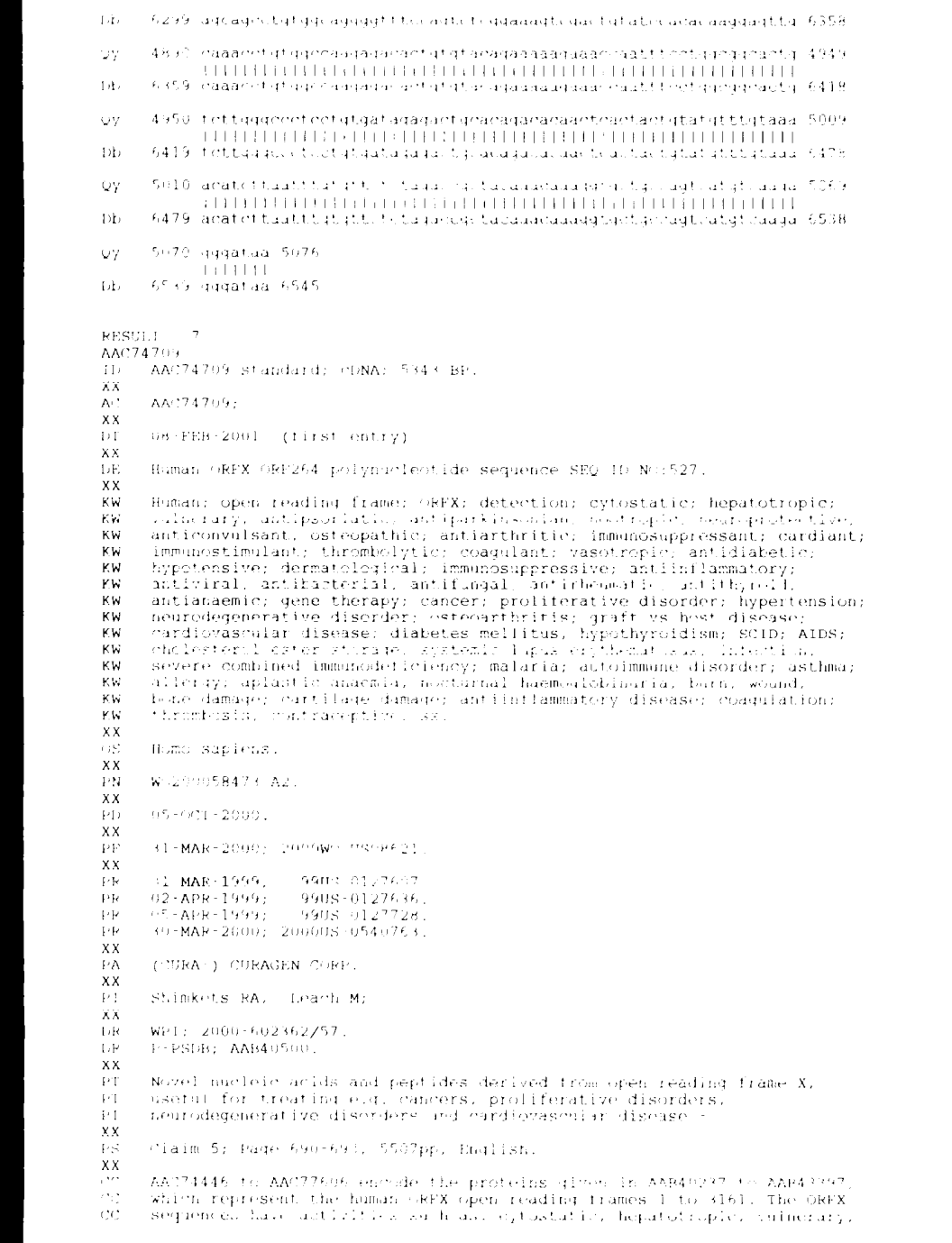
Query Match 41.5%; Score 4755; DB 22; Length 8578;
Best Local Similarity 56.7%; Pred. No. 0;
Matches 5076; Conservative 0; Mismatches 0; Indels 1203; Gaps 4;





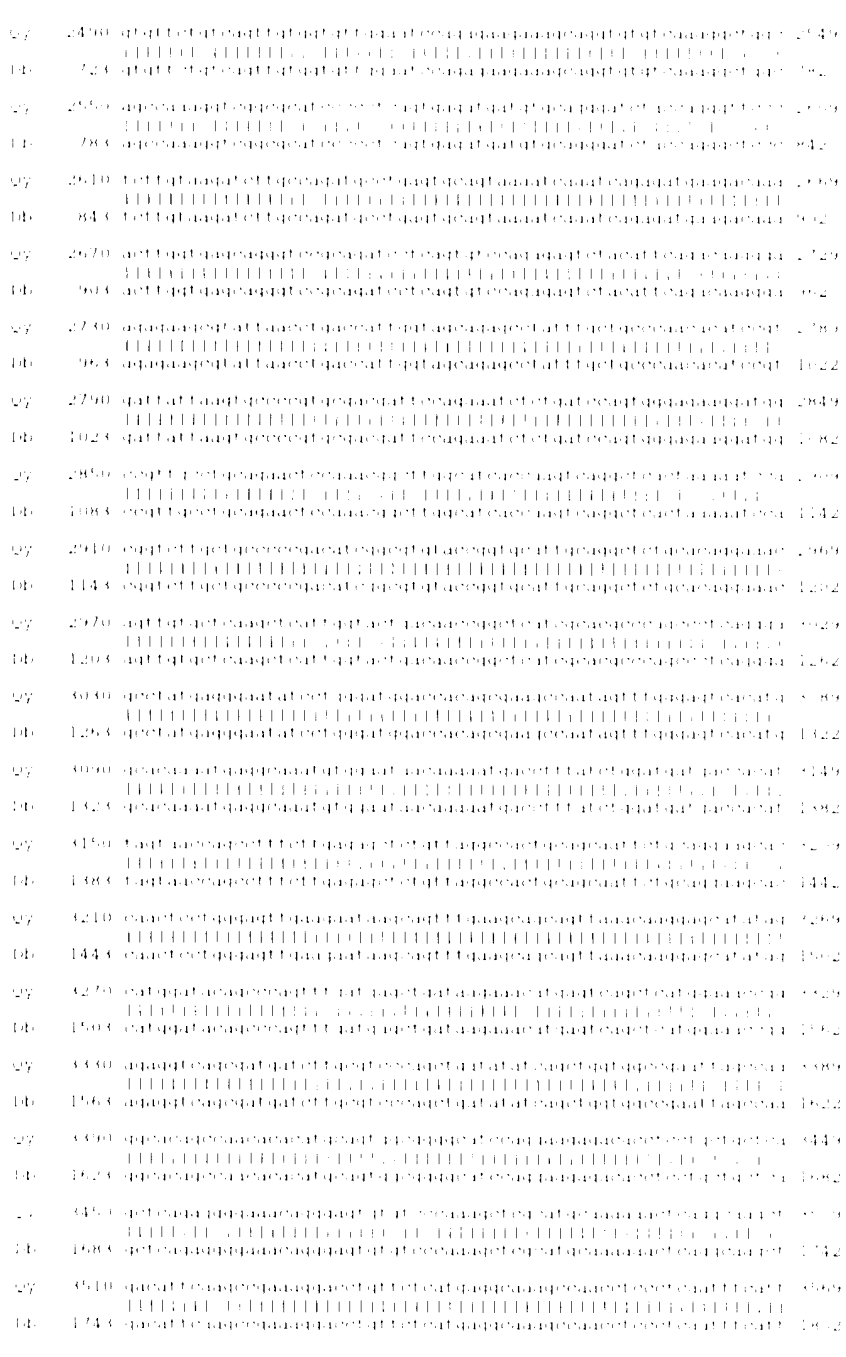


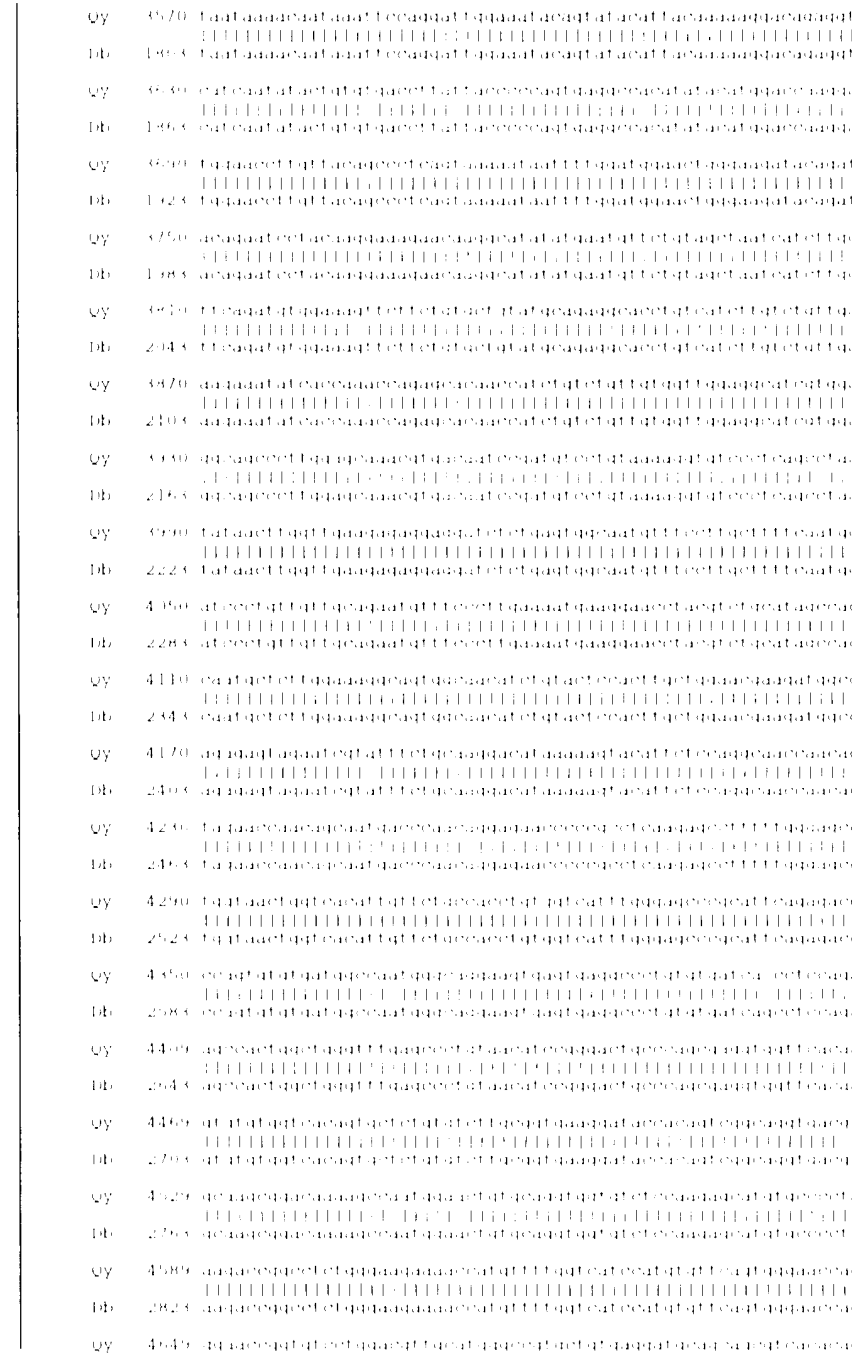




RESULT 7
AAC74709
ID AAC74709 standard; cDNA; 5343 BP.
XX
AC AAC74709;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human ORFX ORF264 polynucleotide sequence SEQ ID NO:527.
XX
KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnerary; antipsoriatic; antiparkinsonian; neurotropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; severe combined
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraception; ss.
XX
OS Homo sapiens.
XX
PN WO200058473-A2.
XX
PD 05-OCT-2000.
XX
PF 31-MAR-2000; 2000WO-US08721.
XX
PR 31-MAR-1999; 99US-0127667.
PR 02-APR-1999; 99US-0127636.
PR 05-APR-1999; 99US-0127728.
PR 30-MAR-2000; 2000US-0540764.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Skiminis RA, Loarh M;
XX
DR WPI; 2000-602362/57.
DR P-PSDB; AAB40500.
XX
PT Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
XX
PS Claim 5; Page 690-691; 5592pp; English.
XX
CC AAC74446 to AAC77696 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have utility in the human cytostatic, hepatotropic, vulnerary,
CC antipsoriatic, antiparkinsonian, neurotropic, neuroprotective,
CC osteopathic, anticonvulsant, antiarthritic, immunosuppressant,
CC immunostimulant, cardiant, thrombolytic, coagulant, vasotropic,
CC antidiabetic, hypotensive, dermatological, immunosuppressive,
CC antiinflammatory, antibacterial, antiviral, antifungal, antirheumatic,
CC antithyroid, and antianaemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORFX-associated disorder. The
CC nucleic acids can be used to express ORFX proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ Sequence 5343 BP; 1496 A; 1137 C; 1304 G; 1404 T; 2 other;

Query Match 64.8%; Score 3289.6; DB 21; Length 5343;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 3304; Conservative 0; Mismatches 4; Indels 1; Gaps 1;





```
Db   2882 ggaacaggtgtacctggacagttgcatgggccgtgctgtgaggtgcagcagcgtcacacag 2942

Qy   4709 cttgtcaacacaacaacagctctgactccaactgtgatgacagaaagagacccacctaaqaa 4768
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2943 cttgtcaacacaacaacagctctgactccaactgtgatgacagaaagagacccacctaaqaa 3002

Qy   4769 qqaactqcacatcaqqqqcctqtgatgtgtgttggcacacaqgccctqqaaqccctgta 4828
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3003 ggaactgcacatcaggggcctgtgatgtgtgttggcacacaggccctqqaaqccctgta 3062

Qy   4829 cagcaqcctqtqqcaqqqtttccaqtctcqqaaaatcqactqtatccacacaaqgaatt 4888
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3063 cagcagcctgtggcagggtttccaqctcqqaaatcgactgtatccacacaaggaatt 3122

Qy   4889 gcaaaccatgtqqcaaaqaaaacatgtgtacaqaaaaagaaaccaatttcctggcggcact 4948
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3123 gcaaaccatgtggcaaagaaaacatgtgtacagaaaaagaaaccaatttcctggcggcact 3182

Qy   4949 gtcttqqqccctcctgtgataqaqactgcacagacacaactcactactgtatgtttgtaa 5008
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3183 gtcttgggccctcctgtgatagagactgcacagacacaactcactactgtatgtttgtaa 3242

Qy   5009 aacatcttaatttgtgtttctctagacctctacaaacaaaqtgtctccaqtcatgtcaaq 5068
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3243 aacatcttaatttgtgtttctctagacctctacaaacaaaqtgtctccaqtcatgtcaaq 3302

Qy   5069 aggataa 5076
          |||||||
Db   3303 aggataa 3310


RESULT   8
AAF87125
ID   AAF87125 standard; DNA; 2912 BP.
XX
AC   AAF87125;
XX
DT   26-MAR-2002  (first entry)
XX
DE   NOV14 coding sequence.
XX
KW   NOV; Cytostatic; contraceptive; antiinflammatory; immunomodulatory;
KW   cardiovascular; casein kinase II phosphorylation site; contraception;
KW   serine/threonine kinase; Peutz Jeghers syndrome; cellular proliferation;
KW   epidermal growth factor; cell development; apoptosis; cell adhesion;
KW   growth migration; cell structure; motility; cancer; immune disorder;
KW   inflammatory disorder; cellular adhesion disorder; long-QT syndrome;
KW   cardiovascular disease; hypertrophic cardiomyopathy; marfan syndrome;
KW   therapy; NOV1; NOV2; NOV3; NOV4; NOV5; NOV6; NOV7; NOV8; NOV9; NOV10;
KW   NOV11; NOV12; NOV13; NOV14; NOV15; NOV16; ds.
XX
OS   Unidentified.
XX
FH   Key             Location/Qualifiers
FT   CDS             98..2878
FT                   /*tag= a
FT                   /product= "NOV14"
XX
PN   WO200136638-A2.
XX
PD   25-MAY-2001.
XX
PF   17-NOV-2000; 2000WO-US31543.
XX
PR   19-NOV-1999;   99US-166336P.
PR   29-NOV-1999;   99US-167785P.
PR   08-MAR-2000;   2000US-187844P.
PR   16-NOV-2000;   2000US-0715417.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Shimkets RA, Lichenstein H, Vernet C, Fernandes E;

XX
DR   WPI; 2001-648134/74.
DR   P-PSDB; AAB83370.
XX
PT   Novel human polypeptides and the nucleic acids that encode them useful
PT   for preventing, diagnosing and treating e.g. cancer, inflammation and
PT   immune disorders -
XX
PS   Claim 9; Page 43-45; 141pp; English.
XX
CC   This sequence encodes the NOV14 protein. The invention relates to
CC   the NOV1-NOV16 proteins, and their coding sequences. The proteins have
CC   cytostatic; contraceptive; antiinflammatory; immunomodulatory; and
CC   cardiovascular activities. The sequences may be used in the prevention,
CC   diagnosis and treatment of diseases associated with inappropriate NOVX
CC   expression. They may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of protein by expressing inactive proteins or to
CC   supplement the patients own production of protein. They are used to
CC   produce NOVX proteins, by inserting the nucleic acid into a cell and
CC   culturing it to express the protein. The DNA may be used as DNA probes in
CC   assays to detect and quantitate the presence of similar DNAs in samples,
CC   and which patients may need restorative therapy. The NOVX protein may
CC   also be used as antigens in the production of antibodies (Abs) against
CC   NOVX and in assays to identify modulators of NOVX expression and
CC   activity. The anti-NOVX Abs and antagonist are used to down regulate
CC   expression and activity. The anti-NOVX Abs are used for detecting the
CC   presence of NOVX in samples. Disorders that may be prevented, diagnosed
CC   and/or treated vary depending on the NOVX protein. NOV1, NOV3, NOV5,
CC   NOV7, NOV9-11 and NOV13-16 have casein kinase II phosphorylation sites
CC   characteristic of serine/threonine kinases, and are used to treat
CC   kinase-related disorders (e.g. Peutz Jeghers syndrome, cellular
CC   proliferation and contraception). NOV2-3, NOV6 and NOV8 are homologous to
CC   the epidermal growth factor (EGF)-like super family and are involved in,
CC   e.g. regulation of cell development, apoptosis, cell adhesion, growth
CC   migration, cell structure and motility and protein management, and are
CC   used to treat cancers, inflammatory disorders, immune disorders and
CC   cellular adhesion disorders. NOV6-10 are homologous to EGF-like
CC   fibrillin proteins and are used to treat cardiovascular disease e.g.
CC   hypertrophic cardiomyopathy, long QT syndrome and marfan syndrome.
XX
SQ   Sequence 2912 BP; 810 A; 713 C; 754 G; 635 T; 0 other;


  Query Match             56.1%;  Score 2862.8;  DB 22;  Length 2912;
  Best Local Similarity   99.8%;  Pred. No. 0;
  Matches 2867;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy    943 cagctttaacagcccccggcgtctttgtcgtagaaaaacacaacagtqqaaatttagaqqqq 992
          ||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db      1 cagctttaacagcccccggcgtctttgtcgtagaaaaacacaacagtqqaaatttagaqqqq 60

Qy    993 ctccgaqaggcaaactttaaqattccaggacctctgatggctgatttcatcttcaaqac 962
          |||||||||||||||||||||||||||||| ||| ||||||||||||||||||||||||
Db      61 ctccgaqaggcaaactttaaqattccaggccctttgatggctgatttcatcttcaaqac 120

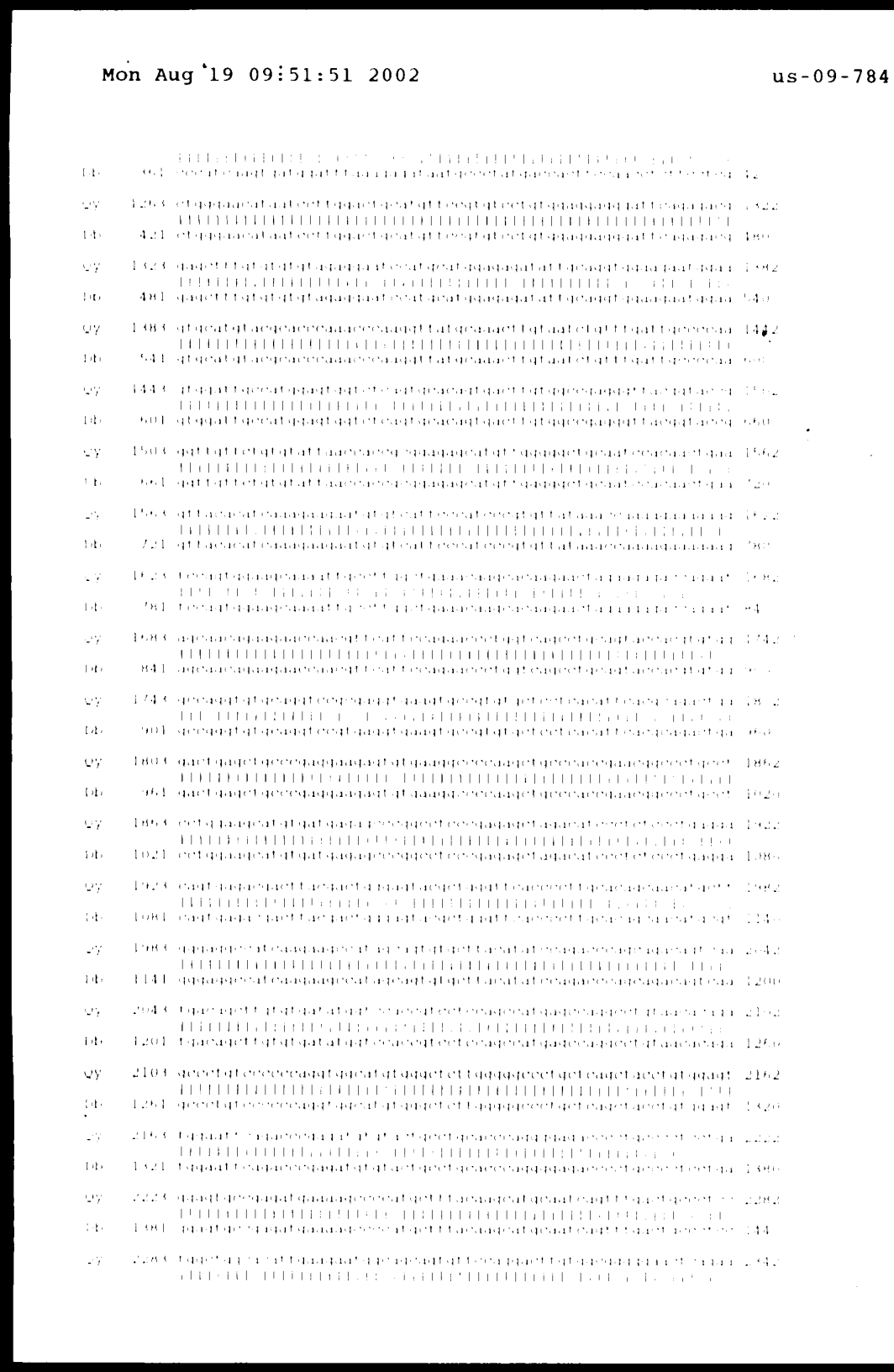
Qy    963 caggtacactqcagccaaaqacagcgtggttcagttcttcttttaccagcccatcaqtca 1022
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 caggtacactqcagccaaaqacagcgtggttcagttcttcttttaccagcccatcaqtca 180

Qy   1023 tcagtggagacaaactqacttctttccctqcactgtgacgtgtggaggagqttatcaqct 1082
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 tcagtggagacaaactqacttctttccctqcactgtgacgtgtggaggagqttatcaqct 240

Qy   1083 caattctgctgaatgtgtggatatccgcttgaagagggtagttcctqaccattattgtca 1142
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 caattctgctgaatgtgtggatatccgcttgaagagggtagttcctqaccattattgtca 300

Qy   1143 ctactaccctgaaaatgtaaaaccaaaaccaaaactgaaggaatgcagcatggatccctq 1202
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 ctactaccctgaaaatgtaaaaccaaaaccaaaactgaaggaatgcagcatggatccctq 360

Qy   1203 gccatcaagtgatggatttaaqqqatgatgcccatgaccactccaacctctcctcg 1262
```



```
Qy    3423 ccaggaaagaatacacct-ctactcttcagctcaaaaaaaaaaaaaaaaqqqaagtgtgtgtcccaaaq 3482
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2581 ccaggaaagaatacacctcctgctgctcagctcaaaaaaaaaaaaaaaqqqaagtgtgtgtcccaaaq 2640

Qy    3483 ctgtgtatgtaaaaaaaatgtgaaaaaaaqtgaacattaaaaaaaaaaaaaaaaaaaaaaaaaaaaaq 3542
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2641 ctgtgtatgtaaaaaaa-tgtgaaaaaaaqtgaacattaaaaaaaaaaaaaaaaaaaaaaaaaaaaaq 2700

Qy    3543 gcaaaagccaaacctcctcaaattcatttaataaaacaataaatttccaggattgqaaaatac 3602
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2701 gcaaaag-caaacctcctcaaattcatttaataaaacaataaatttccaggattgqaaaatac 2760

Qy    3603 agtgtatcattacaaaaaaqtaaagaaattatcaatatactgtatgaacttattaccccaq 3662
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2761 agtgtatcattacaaaaaqtaaagaaattatcaatatactgtatgaacttattaccccaq 2820

Qy    3663 tgaggcaacatatacatgaacaaaqaatgqaaccttgttaaaaaccctaaaa 3716
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2821 tgaggcaacatatacatgaacaaaqaatgqaaccttgttaaaaaccctaaaa 2874


RESULT   9
AAD14369
ID   AAD14369 standard; cDNA; 2538 BP.
XX
AC   AAD14369;
XX
DT   01-NOV-2001  (first entry)
XX
DE   Novel human protein (NHP) cDNA #6.
XX
KW   Novel human protein; NHP; thrombospondin; gene therapy; cancer;
KW   cytostatic; antisense therapy; angiogenesis; biological disorder; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..2538
FT                   /*tag= a
FT                   /product= "Novel human protein"
XX
PN   WO200161011-A2.
XX
PD   23-AUG-2001.
XX
PF   15-FEB-2001; 2001WO-US05290.
XX
PR   17-FEB 2000; 2000US-0184282.
XX
PA   (LEXI ) LEXICON GENETICS INC.
XX
PI   Donoho G, Scoville J, Turner CA, Friedrich G, Zambrowicz B;
PI   Sands AT;
XX
DR   WPI; 2001-514776/56.
DR   P-PSDB; AAE07868.
XX
PT   An isolated nucleic acid encoding a thrombospondin useful as a
PT   hybridization probe and gene therapy treatments of cancer -
XX
PS   Disclosure; Page 42 43; 56pp; English.
XX
CC   The invention relates to novel human polynucleotides encoding proteins
CC   that share sequence similarity with animal proteins having thrombospondin
CC   repeats. NHPs are expressed in human cell lines, pituitary, lymph nodes,
CC   prostrate, testis, adrenal gland, uterus, foetal kidney, foetal lung and
CC   gene trapped human cells. NHPs are thrombospondins useful for treating
CC   biological disorders involving angiogenesis, cancer and development and
CC   also in pharmacogenomic applications. NHPs are useful as a hybridisation
CC   probe for screening libraries, assessing gene expression patterns and
CC   also in gene therapy. Proteins having thrombospondin repeats act as
CC   receptors, secreted extracellular matrix proteins and proteases. The
CC   present sequence is a cDNA encoding novel human protein (NHP).
XX
SQ   Sequence 2538 BP; 670 A; 627 C; 678 G; 563 T; 0 other;


  Query Match             49.1%;  Score 2490;  DB 22;  Length 2538;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2490;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 atgqcttcctqqacaqaqcccctqatqqgtqctqatagqqatqgtcttcatqcactctccc 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 atgqcttcctqqacaqaqcccctqatqqgtqctqatagqqatqgtcttcatqcactctccc 60

Qy      61 ctccqcaqaqcaqcaqaqctqaqaaatctcctqaqaqcctatttcttccccaqatttqcactt 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ctccqcaqaqcaqcaqaqctqaqaaatctcctqaqaqcctatttcttccccaqatttqcactt 120

Qy     121 tctcctcaqqqaaqtttctqaqaaqacaqaaqaqqqaqcaqttcctcactatcqctat 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 tctcctcaqqqaaqtttctqaqaaqacaqaaqaqqqaqcaqttcctcactatcqctat 180

Qy     181 qatqaccaqaqcttcaaqaaaacactqttccaqatqaaqacaaaqatqqcaactqqqatqct 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 qatqaccaqaqcttcaaqaaaacactqttccaqatqaaqacaaaqatqqcaactqqqatqct 240

Qy     241 tqqqqcqactqqaqtqactqctccqqaccctatqqqqqaaqaacatcatattctctqcqq 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 tqqqqcqactqqaqtqactqctccqqaccctatqqqqqaaqaacatcatattctctqcqq 300

Qy     301 aqatqtttqactqqaaqqaattqtqaaqqqcaqaacattcqqtacaaqacatqcaqcaat 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 agatqtttgactggaaggaattgtgaagggcagaacattcggtacaaacatgcagcaat 360

Qy     361 catgactgcgcctccaqatgtcaqtaqatttccaqaqcccaqcagtgctcaqcctacaatqat 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 catgactgcgcctccaqatgtcaqtaqatttccaqaqcccaqcagtgctcaqcctacaatqat 420

Qy     421 gtccaqtatcaqqqqcattactatqaatqqcttccacqatataatqatcctqctqccccq 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 gtccaqtatcaqqqqcattactatqaatqqcttccacqatataatqatcctqctqccccq 480

Qy     481 tqtqcactcaaqtqtcatqcacaaqqacaaaacttqqtqqtqqaqctqqcacctaaqqta 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 tqtqcactcaaqtqtcatqcacaaqqacaaaacttqqtqqtqqaqctqqcacctaaqqta 540

Qy     541 ctqqatqqaactcqttqcaacacqqactcctqqacatatqtatcaqtqccatctqtcaq 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 ctqqatqqaactcqttqcaacacqqactcctqqacatatqtatcaqtqccatctqtcaq 600

Qy     601 qcaqtqqqctqcqatcqqcaactqqaaqcaatqccaaqqaqqacaactqtqqaqtctqt 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 qcaqtqqqctqcqatcqqcaactqqaaqcaatqccaaqqaqqacaactqtqqaqtctqt 660

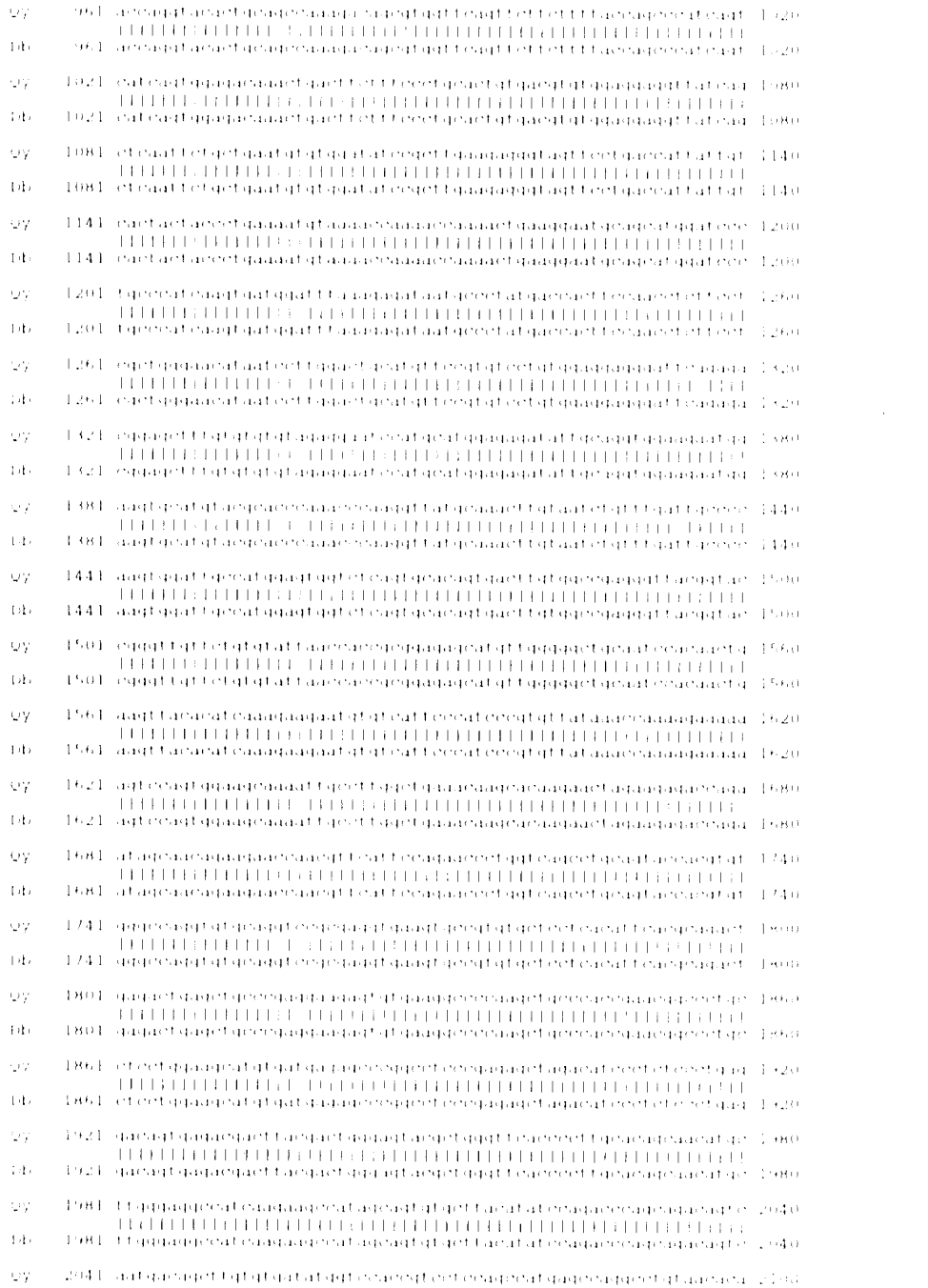
Qy     661 qccqqcqatqqctcaactqcaqqctttqtacaqqqacaatcaaaqtcacacqttctcct 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 qccqqcqatqqctcaactqcaqqctttqtacaqqqacaatcaaaqtcacacqttctcct 720

Qy     721 gaaaaaaqagaaqaaaatgtaattqctqttcctttgqaaaqtcgaaqtqtqaqaattaca 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 gaaaaaaqagaaqaaaatgtaattqctqttcctttgqaaaqtcgaaqtqtqaqaattaca 780

Qy     781 gtqaaaqqaccqtqccacctctttattgaatcaaaaacacttcaaqaaaqcaaaqaaaa 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 gtgaaaqgaccgtgccacctctttattgaatcaaaaacacttcaaqaaaqcaaaqgaaaa 840

Qy     841 ccaqcttttaaccaqcccqqcqtcttqtqtqtaaaaaacaaqcaqtqaaatttcaqaaq 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 caaqcttttaaccaqcccqqcqtcttqtqtqtaqaaaaacaaqcaqtqaaatttcaqaaq 900

Qy     901 ggctcqqaqaggcaaactttaaqattccaaqacctqatqqctqatttcatcttcaaq 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 ggctcqqaqaggcaaactttaaqattccaaqacctqatqqctqatttcatcttcaaq 960
```



RESULT 10
AAD14370
ID AAD14370 standard; cDNA; 2316 BP.
XX
AC AAD14370;
XX
DT 01-NOV-2001 (first entry)
XX
DE Novel human protein (NHP) cDNA #7.
XX
KW Novel human protein; NHP; thrombospondin; gene therapy; cancer;
KW cytostatic; antisense therapy; angiogenesis; biological disorder; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..2316
FT /*tag= a
FT /product= "Novel human protein"
XX
PN WO200161011-A2.
XX
PD 23-AUG-2001.
XX
PF 15-FEB-2001; 2001WO-US05290.
XX
PR 17-FEB-2000; 2000US-0183282.
XX
PA (LEXI-) LEXICON GENETICS INC.
XX
PI Donoho G, Scoville J, Turner CA, Friedrich G, Zambrowicz B;
PI Sands AT;
XX
DR WPI; 2001-514776/56.
DR P-PSDB; AAE07869.
XX
PT An isolated nucleic acid encoding a thrombospondin useful as a
PT hybridization probe and gene therapy treatments of cancer
XX
PS Disclosure; Page 45-46; 56pp; English.
XX

```
CC   The invention relates to novel human polynucleotides encoding proteins
CC   that share sequence similarity with animal proteins having thrombospondin
CC   repeats. NHPs are expressed in human cell lines, pituitary, lymph nodes,
CC   prostrate, testis, adrenal gland, uterus, foetal kidney, foetal lung and
CC   gene trapped human cells. NHPs are thrombospondins useful for treating
CC   biological disorders involving angiogenesis, cancer and development and
CC   also in pharmacogenomic applications. NHPs are useful as a hybridisation
CC   probe for screening libraries, assessing gene expression patterns and
CC   also in gene therapy. Proteins having thrombospondin repeats act as
CC   receptors, secreted extracellular matrix proteins and proteases. The
CC   present sequence is a cDNA encoding novel human protein (NHP).
XX
SQ   Sequence 2316 BP; 617 A; 575 C; 616 G; 508 T; 0 other;

  Query Match             42.9%;  Score 2180;  DB 22,  Length 2316;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2180;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     311 ctggaaggaattgtgaagggcagaacattcggtacaagacatgcagcaatcatgactgcc 370
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      89 ctggaaggaattgtgaagggcagaacattcggtacaagacatgcagcaatcatgactgcc 148

Qy     371 ctccagatgcagaagatttcagagccccagcagtgctcagcctacaatgatgtccagtatc 430
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     149 ctccagatgcagaagatttcagagccccagcagtgctcagcctacaatgatgtccagtatc 208

Qy     431 agggcattactatgaatggcttccacgatataatgatcctgctgcccctgtgcactca 490
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     209 agggcattactatgaatggcttccacgatataatgatcctgctgcccctgtgcactca 268

Qy     491 agtgtcatgcacaaggacaaaacttggtggtggagctggcacctaaggtactggatggaa 550
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     269 agtgtcatgcacaaggacaaaacttggtggtggagctggcacctaaggtactggatggaa 328

Qy     551 ctcgttgcaaaaaggaatcctttgaacatgtgtatccagtggcatctgtcaggcagtgggct 610
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     329 ctcgttgcaaaaaggaatcctttgaacatgtgtatccagtggcatctgtcaggcagtgggct 388

Qy     611 gcgatcggcaactgggaagcaatgccaaggaggacaactgtggagtctgtgccggcgatg 670
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     389 gcgatcggcaactgggaagcaatgccaaggaggacaactgtggagtctgtgccggcgatg 448

Qy     671 gctccacctgcaggcttgtacgggacaatcaaagtcacacgtttctcctgaaaaaagag 730
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     449 gctccacctgcaggcttgtacgggacaatcaaagtcacacgtttctcctgaaaaaagag 508

Qy     731 aagaaaatgtaattgctgttccttgggaagtcgaagtgtgaagaattacagtgaaaggac 790
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     509 aagaaaatgtaattgctgttccttgggaagtcgaagtgtgaagaattacagtgaaaggac 568

Qy     791 ctgcccacctctttattgaatcaaaaacacttcaaggaagcaaaggaagaacacagcttta 850
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     569 ctgcccacctctttattgaatcaaaaacacttcaaggaagcaaaggaagaacacagcttta 628

Qy     851 acagccccggcgtctttgtcgtagaaaacacaacagtggaatttcagaggggctccgaga 910
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     629 acagccccggcgtctttgtcgtagaaaacacaacagtggaatttcagaggggctccgaga 688

Qy     911 ggcaaacttttaagattccaggacctctgatggctgatttcatcttcaagaccaggtaca 970
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     689 ggcaaacttttaagattccaggacctctgatggctgatttcatcttcaagaccaggtaca 748

Qy     971 ctgcagccaaagacagcgtggttcagttcttctttaccagcccatcagtcatcagtgga 1030
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     749 ctgcagccaaagacagcgtggttcagttcttctttaccagcccatcagtcatcagtgga 808

Qy    1031 gacaaactgacttctttccctgcactgtgacgtgtggaggaggttatcagctcaattctg 1090
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     809 gacaaactgacttctttccctgcactgtgacgtgtggaggaggttatcagctcaattctg 868

Qy    1091 ctgaatgtgtggatatccgcttgaagaggtagttcctgaccattattgtcactactacc 1150
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     869 ctgaatgtgtggatatccgcttgaagaggtagttcctgaccattattgtcactactacc 928

Qy    1151 ctgaaaatgtaaaaccaaaaccaaaactgaaggaatgcagcatggatcctgcccatcaa 1210
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     929 ctgaaaatgtaaaaccaaaaccaaaactgaaggaatgcagcatggatcctgcccatcaa 988

Qy    1211 gtgatggatttaaagagataatgccctatgaccactccaacctcttcctcgctggtaac 1270
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     989 gtgatggatttaaagagataatgccctatgaccactccaacctcttcctcgctggtaac 1048

Qy    1271 ataatccttggactgcatgttccgtgtcctgtggaggagggattcagagacggagctttg 1330
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1049 ataatccttggactgcatgttccgtgtcctgtggaggagggattcagagacggagctttg 1108

Qy    1331 tgtgtgtagaggaatccatgcatggagagatattgcaggtggaagaatggaagtgcatgt 1390
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1109 tgtgtgtagaggaatccatgcatggagagatattgcaggtggaagaatggaagtgcatgt 1168

Qy    1391 acgcacccaaacccaaggttatgcaaacttgtaatctgtttgattgcccccaagtggattg 1450
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1169 acgcacccaaacccaaggttatgcaaacttgtaatctgtttgattgcccccaagtggattg 1228

Qy    1451 ccatggagtggtctcagtgcacagtgacttgtggccgaggagttacggtaccgggttgttc 1510
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1229 ccatggagtggtctcagtgcacagtgacttgtggccgaggagttacggtaccgggttgttc 1288

Qy    1511 tgtgtattaaccaccgccggagagcatgttgcgggctgcaatccacaactgaagttacaca 1570
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1289 tgtgtattaaccaccgccggagagcatgttgcgggctgcaatccacaactgaagttacaca 1348

Qy    1571 tcaaagaagaatgtgtcattcccatcccgtgttataaaccaaaagaaaaaagtccagtgg 1630
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1349 tcaaagaagaatgtgtcattcccatcccgtgttataaaccaaaagaaaaaagtccagtgg 1408

Qy    1631 aagcaaaattgccttggctgaaacaagcacaagaactagaagagaccagaatagcaacag 1690
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1409 aagcaaaattgccttggctgaaacaagcacaagaactagaagagaccagaatagcaacag 1468

Qy    1691 aagaaccaacgttcattccagaaccctggtcagcctgcagtaccacgtgtgggccagatg 1750
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1469 aagaaccaacgttcattccagaaccctggtcagcctgcagtaccacgtgtgggccagatg 1528

Qy    1751 tgcaggtccgcgaggtgaagtgccgtgtgctcctcacattcacgcagactgagactgagc 1810
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1529 tgcaggtccgcgaggtgaagtgccgtgtgctcctcacattcacgcagactgagactgagc 1588

Qy    1811 tgcccgaggaagagtgtgaagcccccaagctgcccacccgaacgcccctgcctcctgcaag 1870
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1589 tgcccgaggaagagtgtgaagcccccaagctgcccacccgaacgcccctgcctcctgcaag 1648

Qy    1871 catgtgatgagagcccggcctcccgagagctagacatccctctccctgaggacagtgaga 1930
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1649 catgtgatgagagcccggcctcccgagagctagacatccctctccctgaggacagtgaga 1708

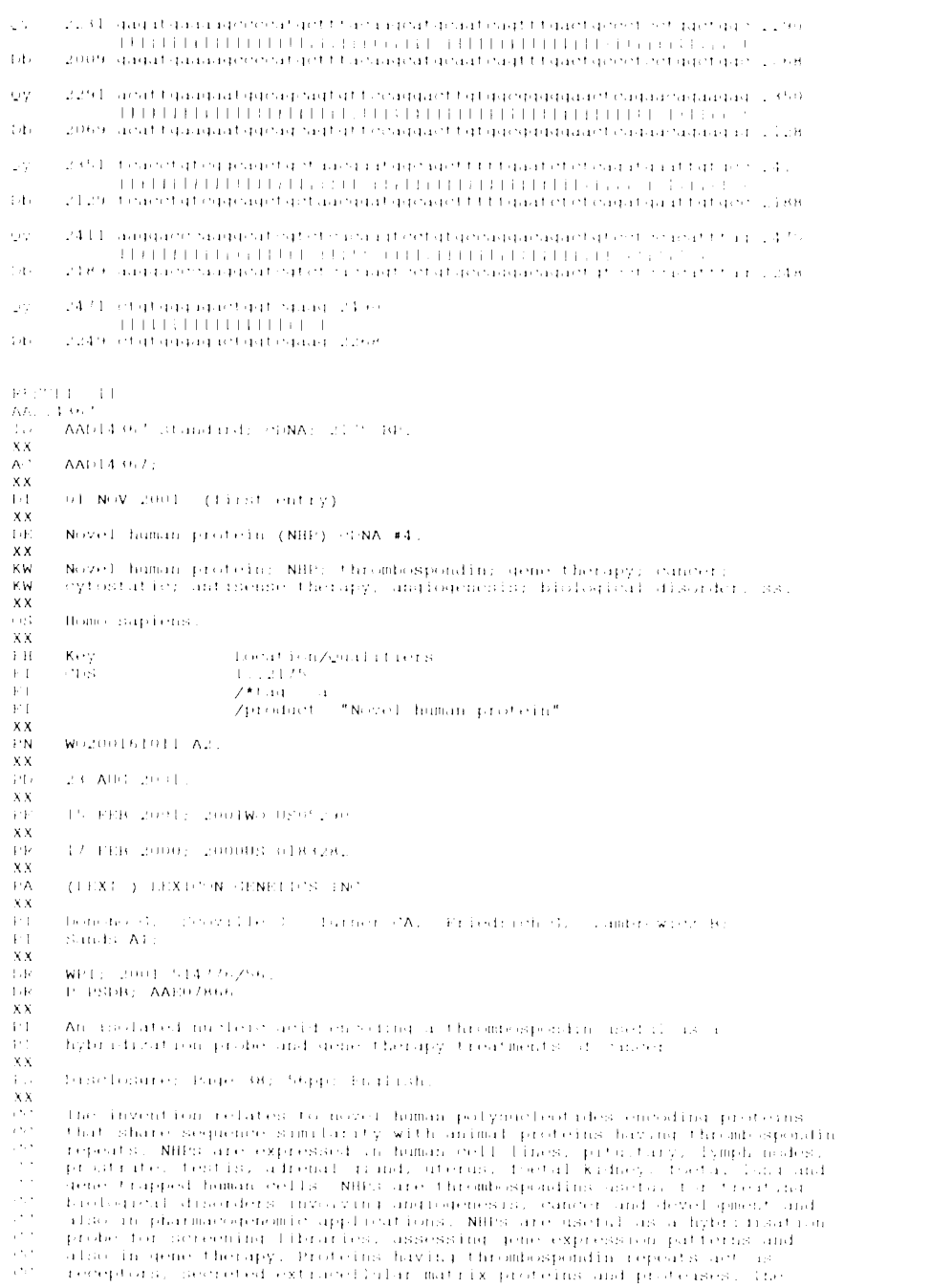
Qy    1931 cgacttacgactgggagtacgctgggttcacccctgcacagcaacatgcttgtgaagccc 1990
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1709 cgacttacgactgggagtacgctgggttcacccctgcacagcaacatgcttgtgaagccc 1768

Qy    1991 atcaagaagccatagcagtgtgcttacatatccagacccagcagacagtcaatgacagct 2050
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1769 atcaagaagccatagcagtgtgcttacatatccagacccagcagacagtcaatgacagct 1828

Qy    2051 tgtgtgatatggtccaccgtcctccagccatgagccaggcctgtaacacagaagccctgtc 2110
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1829 tgtgtgatatggtccaccgtcctccagccatgagccaggcctgtaacacagaagccctgtc 1888

Qy    2111 ccccccaggtggcatgtgggctcttggggggcctgctcagctacctgtggagttggaattc 2170
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1889 ccccccaggtggcatgtgggctcttggggggcctgctcagctacctgtggagttggaattc 1948

Qy    2171 agaccccagatgtgtactgcctgcaccccaggggaaaccccctgccccctcctgaggagtgcc 2230
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1949 agaccccagatgtgtactgcctgcaccccaggggaaaccccctgccccctcctgaggagtgcc 2008
```



cc present sequence is a cDNA encoding novel human protein (NHP).
XX
Sq Sequence 2125 BP; 583 A; 542 C; 575 G; 485 T; 0 other;

Query Match 41.7%; Score 2119; DB 22; Length 2125;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 2126; Conservative 0; Mismatches 10; Indels 0; Gaps 0;



RESULT 12
AAI14368
ID AAB14368 standard; cDNA; 1953 BP.
XX
AC AAB14368;
XX
DT 01-NOV-2001 (first entry)
XX
DE Novel human protein (NHP) cDNA #5.
XX
KW Novel human protein, NHP, thrombospondin, gene therapy, cancer;
KW cytostatic; antisense therapy; angiogenesis; biological disorder; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..1953
FT /*tag= a
FT /product= "Novel human protein"
XX
PN WO200161011-A2.
XX
PD 23-AUG-2001.
XX
PF 15-FEB-2001; 2001WO-US05290.
XX
PR 17-FEB-2000; 2000US-0183282.
XX
PA (GLAX-) GLAXO GROUP LTD.
XX
PI Donche G, Scoville J, Turner CA, Friedrich G, Zambrowicz B;
PI Sands AT;
XX
DR WPI; 2001-514776/56.
DR P-PSDB; AAB07867.
XX
PT An isolated nucleic acid encoding a thrombospondin useful as a
PT hybridization probe and gene therapy treatments of cancer -
XX
PS Disclosure; Page 40-41; 56pp; English.
XX
CC The invention relates to novel human polynucleotides encoding proteins
CC that share sequence similarity with animal proteins having thrombospondin
CC repeats. NHPs are expressed in human cell lines, pituitary, lymph nodes,
CC prostate, testis, adrenal gland, uterus, foetal kidney, foetal lung and
CC gene trapped human cells. NHPs are thrombospondins useful for treating
CC biological disorders involving angiogenesis, cancer and development and
CC also in pharmacogenomic applications. NHPs are useful as a hybridisation
CC probe for screening libraries, assessing gene expression patterns and
CC also in gene therapy. Proteins having thrombospondin repeats act as
CC receptors, secreted extracellular matrix proteins and proteases. The
CC present sequence is a cDNA encoding novel human protein (NHP).
XX
SQ Sequence 1953 BP; 530 A; 480 C; 513 G; 430 T; 0 other;

Query Match 45.6%; Score 1809; DB 22; Length 1953;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 1815; Conservative 0; Mismatches 10; Indels 0; Gaps 0;



RESULT 13
AAD14665
ID AAD14665 standard; cDNA; 1341 BP.
XX
AC AAD14665;
XX
DT 01-NOV-2001 (first entry)
XX
DE Novel human protein (NHP) cDNA #2.
XX
KW Novel human protein; NHP; thrombospondin; motif therapy; cancer;
KW cytostatic; antisense therapy; angiogenesis; bifunctional disorder; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..1341
FT /*tag= a
FT /product= "Novel human protein."
XX
PN WO200164737-A2.
XX
PD 07-SEP-2001.
XX
PF 15-FEB-2001; 2001WO-US05226.
XX

```
PR   17-FEB-2000; 2000US-0183282.
XX
PA   (LEXI-) LEXICON GENETICS INC.
XX
PI   Donoho G, Scoville J, Turner CA, Friedrich G, Zambrowicz B;
PI   Sands AT;
XX
DR   WPI; 2001-514776/56.
DR   P-PSDB; AAE07864.
XX
PI   An isolated nucleic acid encoding a thrombospondin useful as a
PI   hybridization probe and gene therapy treatments of cancer -
XX
PS   Disclosure; Page 95; 256pp; English.
XX
CC   The invention relates to novel human polynucleotides encoding proteins
CC   that share sequence similarity with animal proteins having thrombospondin
CC   repeats. NHPs are expressed in human cell lines, pituitary, lymph nodes,
CC   prostrate, testis, adrenal gland, uterus, foetal kidney, foetal lung and
CC   gene trapped human cells. NHPs are thrombospondins useful for treating
CC   biological disorders involving angiogenesis, cancer and development and
CC   also in pharmacogenomic applications. NHPs are useful as a hybridisation
CC   probe for screening libraries, assessing gene expression patterns and
CC   also in gene therapy. Proteins having thrombospondin repeats act as
CC   receptors, secreted extracellular matrix proteins and proteases. The
CC   present sequence is a cDNA encoding novel human protein (NHP).
XX
SQ   Sequence 1341 BP; 427 A; 315 C; 349 G; 220 T; 0 other;
```

```
  Query Match          28.1%;  Score 1427;  DB 22;  Length 1341;
  Best Local Similarity   99.2%;  Pred. No. 0;
  Matches 1339;  Conservative    0;  Mismatches    0;  Indels    2;  Gaps    1;

Qy       1 atggcttcctgaacgaacccctgatgagtgctgataagaatggtcttcatgcactctccc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 atggcttcctgaacgaacccctgatgagtgctgataggaatggtcttcatgcactctccc 60

Qy      61 ctccagcagaccacagctgaagaaatctccatgaagcctatttccttccagaatttgcactt 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ctcccagcagaccacagctgaagaaatctccatgaagcctatttccttccagaatttgcactt 120

Qy     121 tctcctcagggaaagttttctgaaagacacaacagaggagcagttcctcacttatcgctat 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 tctcctcagggaaagttttctgaaagacacaacagaggagcagttcctcacttatcgctat 180

Qy     181 gatgaccagacctcaagaaacactcgttcagatgaagacaaagatggcaactcgaatgct 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 gatgaccagacctcaagaaacactcgttcagatgaagacaaagatggcaactcgaatgct 240

Qy     241 tgaggcgactgaagtgactgctgcaggacctgtgaggaggaggcatcatattctctgcgg 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 tgaggcgactgaagtgactgctgcaggacctgtgaggaggaggcatcatattctctgcgg 300

Qy     301 agatgtttgactgaaagaagattgtgaaggcagaaacattcggtacaagaacatgcaggaat 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 agatgtttgactgaaagaagattgtgaaggcagaaacattcggtacaagaacatgcaggaat 360

Qy     361 catgactgccgcctcaggatgccagaagatttcagaagcccagcagtgctcgagctacaatgat 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 catgactgccgcctcaggatgccagaagatttcagaagcccagcagtgctcgagctacaatgat 420

Qy     421 gtccagtatcagaggcattactatgaatggcttccacgatataatgatgctgctgccccg 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 gtccagtatcagaggcattactatgaatggcttccacgatataatgatgctgctgccccg 480

Qy     481 tgtgcactcaagtgtcatgcacaagaacaaaacttggtgttgaagctgccacctaaggta 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 tgtgcactcaagtgtcatgcacaagaacaaaacttggtgttgaagctgccacctaaggta 540

Qy     541 ctggatgaaactcgttgcaacacgaactccttggacatgctgtatcagtggcatctgtcaag 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 ctggatgaaactcgttgcaacacgaactccttggacatgctgtatcagtggcatctgtcaag 600

Qy     601 gcagtggactgcgatgaagactgtggaggcaatgccaaggaggccatgctgtggagtctgt 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 gcagtggactgcgatgaagactgtggaggcaatgccaaggaggccatgctgtggagtctgt 660

Qy     661 gccgagcgatggatccacctgcaaggcttgtaccggggacatcaaaatcaacatttctcct 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 gccgagcgatggatccacctgcaggcttgtaccggggacatcaaaatcaacatttctcct 720

Qy     721 gaaaaaagaagaaagaaaatgtaattgctgttcctttgaggaagtcgaagttgtgaagaattaca 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 gaaaaaagaagaaagaaaatgtaattgctgttcctttgaggaagtcgaagttgtgaagaattaca 780

Qy     781 gtgaaaggacctgccaacctctttattgaatccaaaaacacttcaaggaaagcaaaggaagaa 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 gtgaaaggacctgccaacctctttattgaatccaaaaacacttcaaggaaagcaaaggaagaa 840

Qy     841 cacagctttaacagccccgcgtctgtgtcgtaaaaacacaacaagtggaatttcaagagg 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 cacagctttaacagccccgcgtctgtgtcgtaaaaacacaacaagtggaatttcaagagg 900

Qy     901 agctccgagagaggcaaacttttaagattccagcacctctgatggctgattttcatcttcaag 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 agctccgagagaggcaaacttttaagattccagcacctctgatggctgattttcatcttcaag 960

Qy     961 acagagtacactgcagctgaagacagcgtgattcagttcttctttaccgacccatcagt 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 acagagtacactgcagctgaagacagcgtgattcagttcttctttaccgacccatcagt 1020

Qy    1021 catcagtgaagacagagctgacttcttccctgcactgtgacgtgtgaagaagttatcaag 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 catcagtgaagacagagctgacttcttccctgcactgtgacgtgtgaagaagttatcaag 1080

Qy    1081 ctcaattctgctgaatgtgtggatatccgcttgaagaagggtagttcctgaccattattgt 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 ctcaattctgctgaatgtgtggatatccgcttgaagaagggtagttcctgaccattattgt 1140

Qy    1141 cactactgcctgaaaatgtgaaaacaaaaacaaaactgaaggaatgcagcatggatcc 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 cactactgcctgaaaatgtgaaaacaaaaacaaaactgaaggaatgcagcatggatcc 1200

Qy    1201 tgcccatcaagtgatggatttaaagagataatgccctatgaccacttccaacctcttcct 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 tgcccatcaagtgatggatttaaagagataatgccctatgaccacttccaacctcttcct 1260

Qy    1261 c--gctgaggaacataatccttggactgcatgttccgtgtcctgtgaaggaagaattcaga 1318
           |  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1261 cgaagctggaaacataatccttggactgcatgttccgtgtcctgtgaaggaagaattcaga 1320

Qy    1319 gacgaagctttgtgtgtgtgtag 1339
           ||||||||||||||||||||||
Db    1321 gacgaagctttgtgtgtgtgtag 1341
```

```
RESULT  14
AAD14366
ID   AAD14366 standard; cDNA; 1119 BP.
XX
AC   AAD14366;
XX
DT   01-NOV-2001 (first entry)
XX
DE   Novel human protein (NHP) cDNA #3.
XX
KW   Novel human protein; NHP; thrombospondin; gene therapy; cancer;
KW   cytostatic; antisense therapy; angiogenesis; biological disorder; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1119
FT                   /tag= a
FT                   /product= "Novel human protein"
```

XX
PN Wo200161011-A2.
XX
PD 23-AUG-2001.
XX
PF 15-FEB-2001; 2001WO-US05299.
XX
PR 17-FEB-2000; 2000US-0506202.
XX
PA (HEXI) HEXION GENETICS INC.
XX
PI Benedict C, Scarlett L, Turner A, Friedrich G, Zambrowicz B;
PI Sands AT;
XX
DR WPI; 2001-514776/56.
DR P-PSDB; AAE02865.
XX
PT An isolated nucleic acid encoding a thrombospondin useful as a
PT hybridization probe and gene therapy treatments of cancer
XX
PS Disclosure; Page 96; 59pp; English.
XX
CC The invention relates to novel human polynucleotides encoding proteins
CC that share sequence similarity with animal proteins having thrombospondin
CC repeats. NHPs are expressed in human cell lines, pituitary, lymph nodes,
CC prostrate, testis, adrenal gland, uterus, foetal kidney, foetal liver and
CC gene trapped human cells. NHPs are thrombospondins useful for treating
CC biological disorders involving angiogenesis, cancer and development and
CC also in pharmacogenomic applications. NHPs are useful as a hybridisation
CC probe for screening libraries, assessing gene expression patterns and
CC also in gene therapy. Proteins having thrombospondin repeats act as
CC receptors, secreted extracellular matrix proteins and proteases. The
CC present sequence is a cDNA encoding novel human protein. (NHP).
XX
SQ Sequence 1129 BP; 304 A; 295 C; 282 G; 205 T; 0 other;

Query Match 20.0%; Score 1017; DB 22; Length 1129;
Best Local Similarity 99.8%; Pred. No. 8.2e-92;
Matches 1029; Conservative 0; Mismatches 0; Indels 2; Gaps 1;

Qy 311 ctgaaaggaattgtgaaaggaaaaaattgtgatacaaaaaatgaaaaaaaatgaaaaaaaa 370
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 89 ctgaaaggaattgtgaaaggaaaaaattgtgatacaaaaaatgaaaaaaaatgaaaaaaaa 148

Qy 371 ctccagatgcaggaaaatttcaggaccccagcagtgctgcaccatgatgtgccatatgaaaaa 430
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 149 ctccagatgcaggaaaatttcaggaccccagcagtgctgcaccatgatgtgccatatgaaaaa 208

Qy 431 aggggcatgactatgaatggcttccacgatataatgatgtcctgtgcccctgtgcgctca 490
|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 209 aggggcatgactatgaatggcttccacgatataatgatgtcctgtgcccctgtgcgctca 268

Qy 491 aatatgtatgcgcaaagcaaaactgatgatgaagctgcgacctaaagtactgaaatgaaa 550
|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 269 aatgtgtatgcgcaaagcaaaactgatgatgaagctgcgacctaaagtactgaaatgaaa 328

Qy 551 ctgattgcaaacacgaactcctgcgaacatgtatatcatgaatgcatgtcaaggcatgtgaggcatgagt 610
|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 329 ctgattgcaaacacgaactcctgcgaacatgtatatcatgaatgcatgtcaaggcatgtgaggcatgagt 388

Qy 611 gcgatgcgcaaactgcgaaaccatgcaaaaaaaagcaactgtgaatgtgtgcgcgaatg 670
|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 389 gcgatgcgcaactgcgaaaccatgcaaaaaaaagcaactgtgaatgtgtgcgcgaatg 448

Qy 671 gctccaacctgcaggctgtacgaaaaaaatcaaaatgacacgttctcctgaaaaaaaa 730
|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 449 gctccaacctgcaggctgtacgaaaaaaatcaaaatgacacgttctcctgaaaaaaaa 508

Qy 731 aaaaaaatgtaattgctgttcctttgaaaatgcaaatgtgaaatgtgaaattacaatgaaaaaa 790
|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 509 aaaaaaatgtaattgctgttcctttgaaaatgcaaatgtgaaatgtgaaattacaatgaaaaaa 568

Qy 791 ctgaccctcctctttattgaatgaaaaactttaaggaaaaaaaaaaaaacagctttta 850
|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 569 ctgaccctcctctttattgaatgaaaaactttaaggaaaaaaaaaaaaacagctttta 628

Qy 851 acagcccccgcgtctttgtcgtgtaaaaacacaacaatgaaaatttaaaaaaggctccgaaa 910
|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 629 acagcccccgcgtctttgtcgtgtaaaaacacaacaatgaaaatttaaaaaaggctccgaaa 688

Qy 911 gcgaaacttttaaaattcaaaaactctataactctatttaatttaaaaacaaaaatacca 970
|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 689 gcgaaacttttaaaattcaaaaactctataactctatttaatttaaaaacaaaaatacca 748

Qy 971 ctaaagcaaaaaaaaaagctgaatttcattctttttt aaaaaaaataaataaaaaaaaaa 1030
|||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 749 ctaaagcaaaaaaaaaagctgaatttcattctttttttaaaaaaaataaataaaaaaaaaa 808

Qy 1031 aaaaaactgaattcttttcccctgaactgtgaactgtgaaaaaaattatcaaatcaattctg 1090
|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 809 aaaaaactgaattcttttcccctgaactgtgaactgtgaaaaaaattatcaaatcaattctg 868

Qy 1091 ctaaatgtgtgaaatatccagcttaaaaaaaatgattcctgacaccattattgtcaactactacc 1150
|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 869 ctaaatgtgtgaaatatccagcttaaaaaaatgattcctgacaccattattgtcaactactacc 928

Qy 1151 ctgaaaatgtaaaaccaaaaccaaaactaaaaaaatgcaaccatgaatcctccatccaaa 1210
|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 929 ctgaaaatgtaaaaccaaaaccaaaactaaaaaaatgcaaccatgaatcctccatccaaa 988

Qy 1211 gtgatgaaatttaaaaaaataaatgccctatgaccacttccaaccttctcctc qctgaa 1268
|| ||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db 989 gtgatgaaatttaaaaaaataatgccctatgaccacttccaaccttctcctcaaactgaa 1048

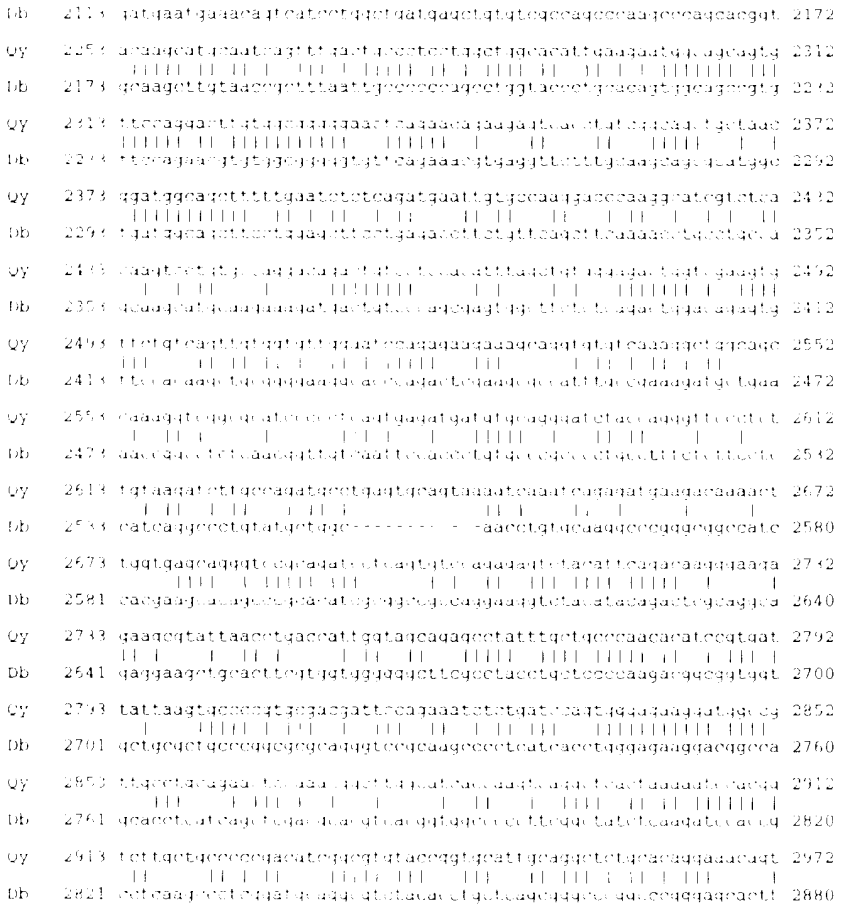
Qy 1269 acataatcctgaaactgcatgttcatgtcctgtgaaaaaaaaaattcaaaacaaaaactt 1328
|||||||||||||||||||||| |||||||||||||||||||||| |||| ||||||||||
Db 1049 acataatcctgaactgcatgttcatgtcctgtgaaaaaaaaaattcaaaacaaaaactt 1108

Qy 1329 tgtgtgtgtaa 1339
|||||||||||
Db 1109 tgtgtgtgtaa 1119

RESULT 15
ABA06468
ID ABA06468 standard; cDNA; 5269 BP.
XX
AC ABA06468;
XX
DT 10-JAN-2002 (first entry)
XX
DE Human cDNA SEQ ID NO: 134.
XX
KW Human; gene therapy; neural disorder; immune system disorder;
KW muscular disorder; reproductive disorder; gastrointestinal disorder;
KW pulmonary disorder; cardiovascular disorder; renal disorder;
KW proliferative disorder; inflammation; ss.
XX
OS Homo sapiens.
XX
PN WO200154474-A2.
XX
PD 26-JUL-2001.
XX
PF 22-JAN-2001; 2001WO-US02349.
XX
PR 31-JAN-2000; 2000US-179067P.
PR 04-FEB-2000; 2000US-180628P.
PR 24-FEB-2000; 2000US-184644P.
PR 02-MAR-2000; 2000US-186450P.
PR 16-MAR-2000; 2000US-189874P.
PR 17-MAR-2000; 2000US-190076P.
PR 18-APR-2000; 2000US-198123P.
PR 19-MAY-2000; 2000US-205515P.
PR 07-JUN-2000; 2000US-209467P.

PR 28-JUN-2000; 2000US-214886P.
PR 30-JUN-2000; 2000US-215145P.
PR 07-JUL-2000; 2000US-216647P.
PR 07-JUL-2000; 2000US-216889P.
PR 11-JUL-2000; 2000US-217487P.
PR 11-JUL-2000; 2000US-217496P.
PR 14-JUL-2000; 2000US-218290P.
PR 26-JUL-2000; 2000US-220963P.
PR 26-JUL-2000; 2000US-220964P.
PR 14-AUG-2000; 2000US-224518P.
PR 14-AUG-2000; 2000US-224519P.
PR 14-AUG-2000; 2000US-225213P.
PR 14-AUG-2000; 2000US-225214P.
PR 14-AUG-2000; 2000US-225266P.
PR 14-AUG-2000; 2000US-225267P.
PR 14-AUG-2000; 2000US-225268P.
PR 14-AUG-2000; 2000US-225270P.
PR 14-AUG-2000; 2000US-225417P.
PR 14-AUG-2000; 2000US-225757P.
PR 14-AUG-2000; 2000US-225758P.
PR 14-AUG-2000; 2000US-225759P.
PR 16-AUG-2000; 2000US-226276P.
PR 22-AUG-2000; 2000US-226681P.
PR 22-AUG-2000; 2000US-226689P.
PR 22-AUG-2000; 2000US-227193P.
PR 23-AUG-2000; 2000US-227009P.
PR 30-AUG-2000; 2000US-229924P.
PR 01-SEP-2000; 2000US-229287P.
PR 01-SEP-2000; 2000US-229414P.
PR 01-SEP-2000; 2000US-229444P.
PR 01-SEP-2000; 2000US-229445P.
PR 05-SEP-2000; 2000US-229509P.
PR 05-SEP-2000; 2000US-229513P.
PR 06-SEP-2000; 2000US-230437P.
PR 06-SEP-2000; 2000US-230438P.
PR 08-SEP-2000; 2000US-231242P.
PR 08-SEP-2000; 2000US-231243P.
PR 08-SEP-2000; 2000US-231244P.
PR 08-SEP-2000; 2000US-231413P.
PR 08-SEP-2000; 2000US-231414P.
PR 08-SEP-2000; 2000US-231780P.
PR 08-SEP-2000; 2000US-231981P.
PR 12-SEP-2000; 2000US-231968P.
PR 14-SEP-2000; 2000US-232467P.
PR 14-SEP-2000; 2000US-232498P.
PR 14-SEP-2000; 2000US-232499P.
PR 14-SEP-2000; 2000US-232500P.
PR 14-SEP-2000; 2000US-232401P.
PR 14-SEP-2000; 2000US-232066P.
PR 14-SEP-2000; 2000US-232064P.
PR 14-SEP-2000; 2000US-232065P.
PR 21-SEP-2000; 2000US-234224P.
PR 21-SEP-2000; 2000US-234274P.
PR 25-SEP-2000; 2000US-234967P.
PR 25-SEP-2000; 2000US-234998P.
PR 26-SEP-2000; 2000US-235184P.
PR 27-SEP-2000; 2000US-235634P.
PR 27-SEP-2000; 2000US-235836P.
PR 29-SEP-2000; 2000US-236427P.
PR 29-SEP-2000; 2000US-236467P.
PR 29-SEP-2000; 2000US-236468P.
PR 29-SEP-2000; 2000US-236536P.
PR 29-SEP-2000; 2000US-236870P.
PR 02-OCT-2000; 2000US-236802P.
PR 02-OCT-2000; 2000US-237037P.
PR 02-OCT-2000; 2000US-237038P.
PR 02-OCT-2000; 2000US-237039P.
PR 02-OCT-2000; 2000US-237040P.
PR 13-OCT-2000; 2000US-239935P.
PR 13-OCT-2000; 2000US-239937P.
PR 20-OCT-2000; 2000US-241660P.
PR 20-OCT-2000; 2000US-241821P.
PR 20-OCT-2000; 2000US-241785P.
PR 20-OCT-2000; 2000US-241786P.
PR 20-OCT-2000; 2000US-241787P.
PR 20-OCT-2000; 2000US-241808P.
PR 20-OCT-2000; 2000US-241809P.
PR 20-OCT-2000; 2000US-241826P.
PR 01-NOV-2000; 2000US-244617P.
PR 08-NOV-2000; 2000US-246474P.
PR 08-NOV-2000; 2000US-246475P.
PR 08-NOV-2000; 2000US-246476P.
PR 08-NOV-2000; 2000US-246477P.
PR 08-NOV-2000; 2000US-246478P.
PR 08-NOV-2000; 2000US-246523P.
PR 08-NOV-2000; 2000US-246524P.
PR 08-NOV-2000; 2000US-246525P.
PR 08-NOV-2000; 2000US-246526P.
PR 08-NOV-2000; 2000US-246527P.
PR 08-NOV-2000; 2000US-246528P.
PR 08-NOV-2000; 2000US-246532P.
PR 09-NOV-2000; 2000US-246609P.
PR 08-NOV-2000; 2000US-246610P.
PR 08-NOV-2000; 2000US-246611P.
PR 08-NOV-2000; 2000US-246613P.
PR 17-NOV-2000; 2000US-249207P.
PR 17-NOV-2000; 2000US-249208P.
PR 17-NOV-2000; 2000US-249209P.
PR 17-NOV-2000; 2000US-249210P.
PR 17-NOV-2000; 2000US-249211P.
PR 17-NOV-2000; 2000US-249212P.
PR 17-NOV-2000; 2000US-249213P.
PR 17-NOV-2000; 2000US-249214P.
PR 17-NOV-2000; 2000US-249215P.
PR 17-NOV-2000; 2000US-249216P.
PR 17-NOV-2000; 2000US-249217P.
PR 17-NOV-2000; 2000US-249218P.
PR 17-NOV-2000; 2000US-249244P.
PR 17-NOV-2000; 2000US-249245P.
PR 17-NOV-2000; 2000US-249264P.
PR 17-NOV-2000; 2000US-249265P.
PR 17-NOV-2000; 2000US-249297P.
PR 17-NOV-2000; 2000US-249299P.
PR 17-NOV-2000; 2000US-249300P.
PR 01-DEC-2000; 2000US-250160P.
PR 01-DEC-2000; 2000US-250391P.
PR 05-DEC-2000; 2000US-251030P.
PR 05-DEC-2000; 2000US-251988P.
PR 05-DEC-2000; 2000US-251719P.
PR 06-DEC-2000; 2000US-251479P.
PR 08-DEC-2000; 2000US-251856P.
PR 08-DEC-2000; 2000US-251868P.
PR 08-DEC-2000; 2000US-251869P.
PR 08-DEC-2000; 2000US-251989P.
PR 08-DEC-2000; 2000US-251990P.
PR 11-DEC-2000; 2000US-251997P.
PR 05-JAN-2001; 2001US-259678P.
XX
PA (HUMA) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WI, 2001-476161/51.
DR P-PSDB; AAB10246.
XX
PT Isolated nucleic acid molecule encoding an inflammation-associated
PT polypeptide is used in preventing, treating or ameliorating a medical
PT condition
XX
PS Claim 1; SEQ ID NO 144; 859pp - Sequence Listing; English.
XX
CC The present invention provides human cDNAs, proteins and related genomic
CC DNAs. These can be used in the treatment of neural, immune system,
CC muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,
CC renal and proliferative disorders and inflammation. The present sequence
CC is a cDNA of the invention.





Search completed: August 18, 2002, 05:55:25
Job time: 10605 sec

SeqSearch version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

GenCore version 4.5: nucleic search, using sw model

Run on: August 18, 2002, 02:55:10 ; Search time 0.64 Seconds
(without alignments)
17696.004 Million cell updates/sec

Title: US-09-784-358-1
Perfect score: 5076
Sequence: 1 atggctctcatgctgcc.....gatgatgtgatttttgtt 507

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1776656 seqs, 1484320864 residues

Total number of hits satisfying chosen parameters: 1776656

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mu:*

First, Nov. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

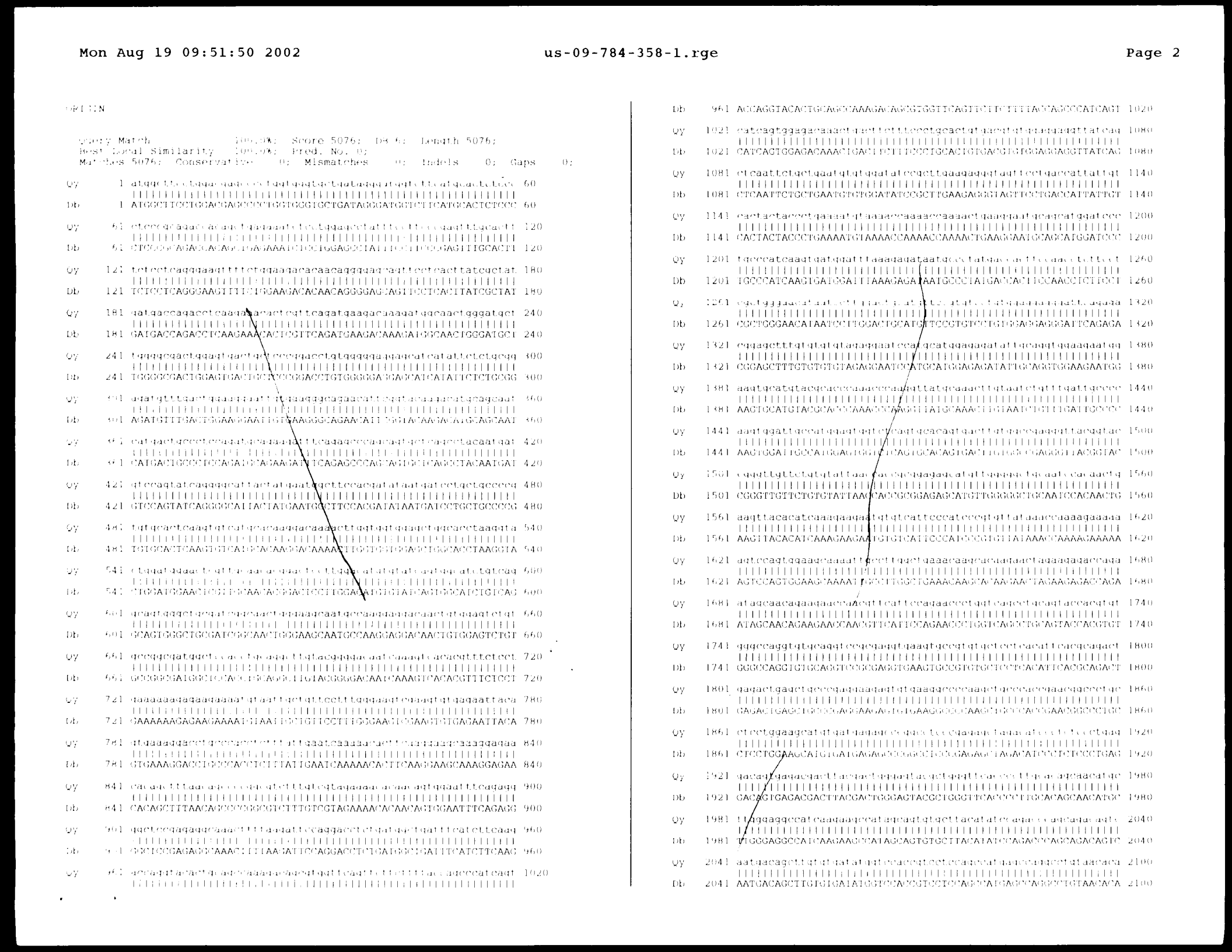
Result No.	% query Score	Match Length	DB	ID	Description
1	5076	100.0	5076	6	AX224809 AX224809 Sequence
2	5066.6	99.8	7265	6	AX235456 AX235456 Sequence
3	4766	93.9	40654	6	AX240828 AX240828 Sequence
4	3862.2	76.1	6503	6	AX235410 AX235410 Sequence
5	3859	76.0	6294	6	AX235454 AX235454 Sequence
6	3755	74.0	8578	6	AX224825 AX224825 Sequence
7	3072.4	60.5	5107	6	AX235437 AX235437 Sequence
8	3072.4	60.5	5107	6	AX235439 AX235439 Sequence
9	3072.4	60.5	5107	6	AX235413 AX235413 Sequence
10	3072.4	60.5	5107	9	AB033059 AB033059 Homo sapi
11	3070.8	60.5	5107	6	AX235411 AX235411 Sequence
12	2895.8	57.0	2912	6	AX149471 AX149471 Sequence
13	2490	49.1	2538	6	AX224819 AX224819 Sequence
14	2295.8	45.2	2528	9	AF277552 AF277552 Homo sapi
15	2185	43.0	2510	6	AX224821 AX224821 Sequence
16	2120.6	41.8	4614	9	AB022736 AB022736 Macaca fa
17	2110	41.7	2175	6	AX224815 AX224815 Sequence
18	1809	35.6	1953	6	AX224817 AX224817 Sequence
19	1527	30.1	1541	6	AX224811 AX224811 Sequence
20	1507.4	29.1	348879	9	AY222994 AY222994 Homo sapi
21	1507.4	29.1	1811627	2	AC027667 AC027667 Homo sapi
22	1484.4	29.4	1859926	2	AC074419 AC074419 Homo sapi
23	1017	20.0	1119	6	AX224813 AX224813 Sequence
24	737.2	14.4	1803	9	AF176613 AF176613 Homo sapi
25	730	14.4	1616	9	AF251858 AF251858 Homo sapi
26	719	14.2	165652	2	AC027605 AC027605 Homo sapi
27	701.2	13.8	169976	2	AC073419 AC073419 Homo sapi
28	698.8	13.8	55279	2	AC110192 AC110192 Mus muscu
29	696.6	13.7	64303	2	AC103754 AC103754 Homo sapi
30	673	13.3	168497	2	AC092410 AC092410 Bos tauru
31	672	13.2	55276	2	AC110192 AC110192 Mus muscu
32	647.4	12.9	214362	2	BX399965 BX399965 Homo sapi
33	641.2	12.8	59982	2	AC103743 AC103743 Homo sapi
34	641.2	12.8	60852	2	AC103750 AC103750 Homo sapi
35	625	12.4	167458	2	AC010673 AC010673 Homo sapi
36	625	12.4	173435	2	AC015715 AC015715 Homo sapi
37	625	12.4	180454	2	AC009689 AC009689 Homo sapi
38	625	12.4	181660	2	AC087738 AC087738 Homo sapi
39	618	12.3	58890	2	AC110050 AC110050 Homo sapi
40	611.4	12.2	2848	9	AF119674 AF119674 Homo sapi
41	604.6	12.1	2145	6	AX327757 AX327757 Sequence
42	604.6	12.1	3132	6	AX327754 AX327754 Sequence
43	604.6	12.1	4199	6	AX111571 AX111571 Sequence
44	604.6	12.1	4207	6	AX327747 AX327747 Sequence
45	604.6	12.1	4312	6	AX198051 AX198051 Sequence

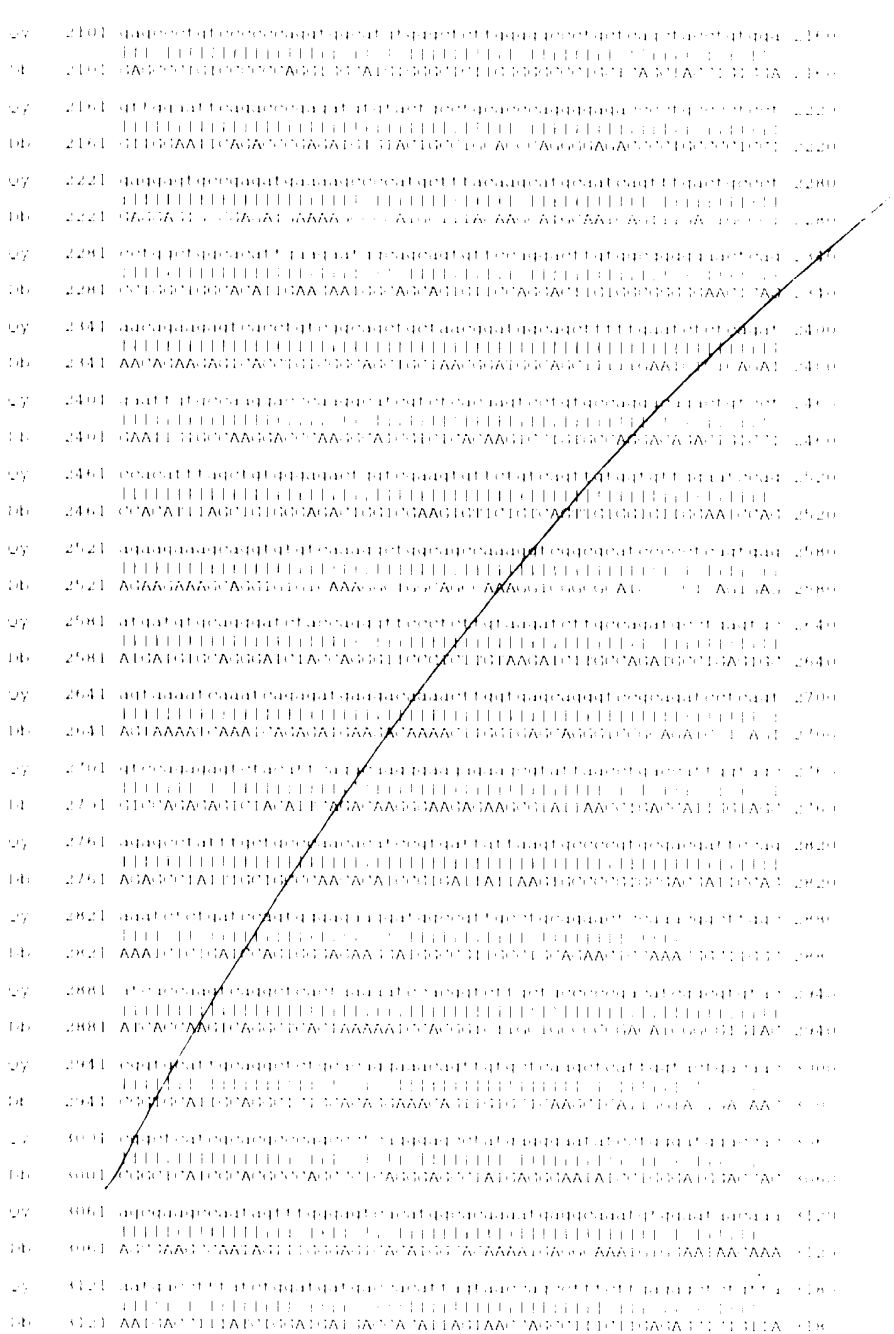
ALIGNMENTS

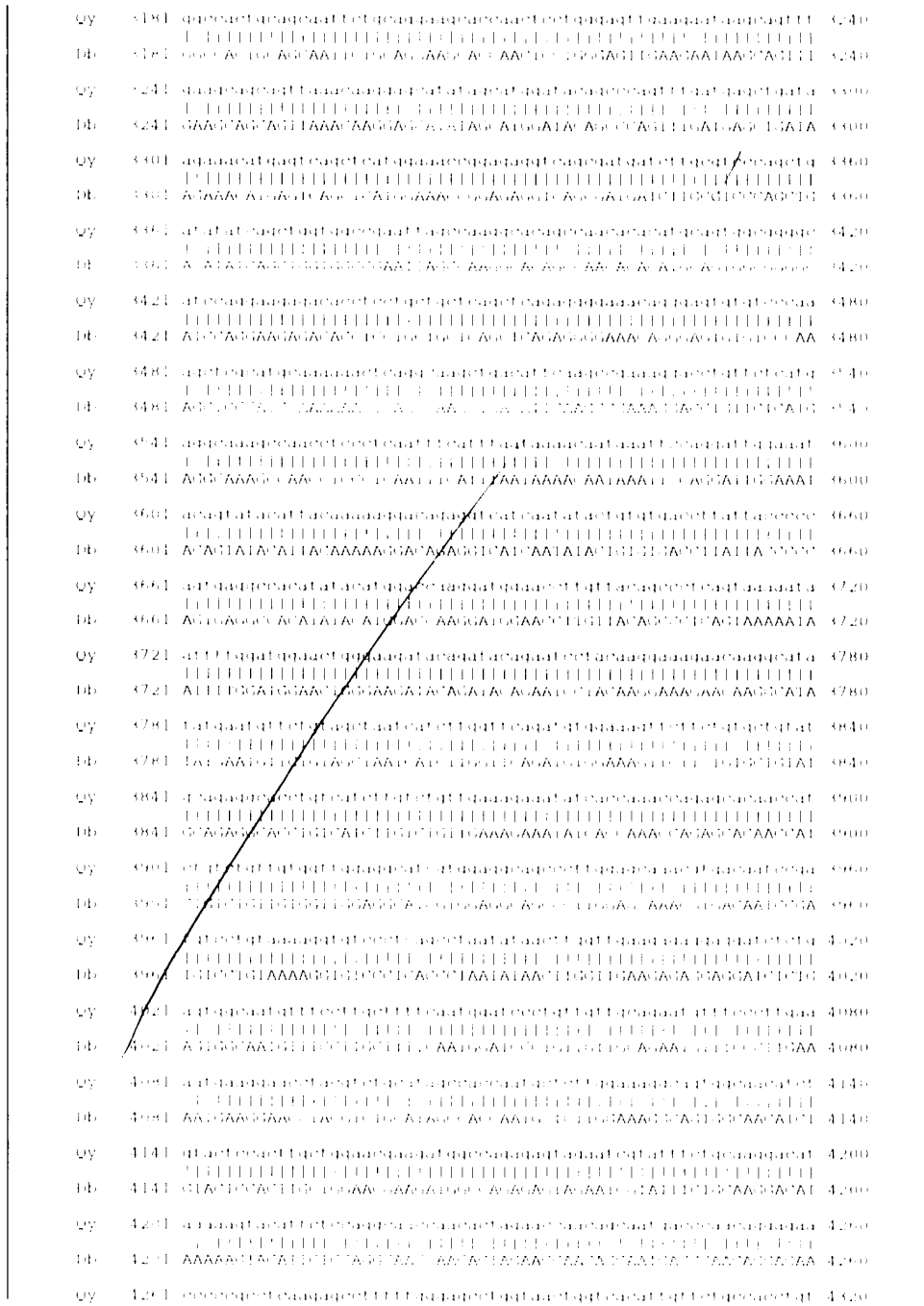
RESULT 1
AX224809
LOCUS AX224809 5076 bp DNA linear PAT 10-SEP-2002
DEFINITION Sequence 1 from Patent WO0161011.
ACCESSION AX224809
VERSION AX224809.1 GI:15564469
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 5076)
AUTHORS Donoho,G., Sexsmith,E., Turner,C.A., Friedrich,G., Zambrowicz,B.
and Sands,A.T.
TITLE Novel human thrombospondin repeat proteins and polynucleotides
encoding the same
JOURNAL Patent: WO 0161011 A 1 23 AUG 2001;
Lexicon Genetics Incorporated (US)
FEATURES Location/Qualifiers
source 1..5076
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 1484 a 1227 c 1230 g 1135 t

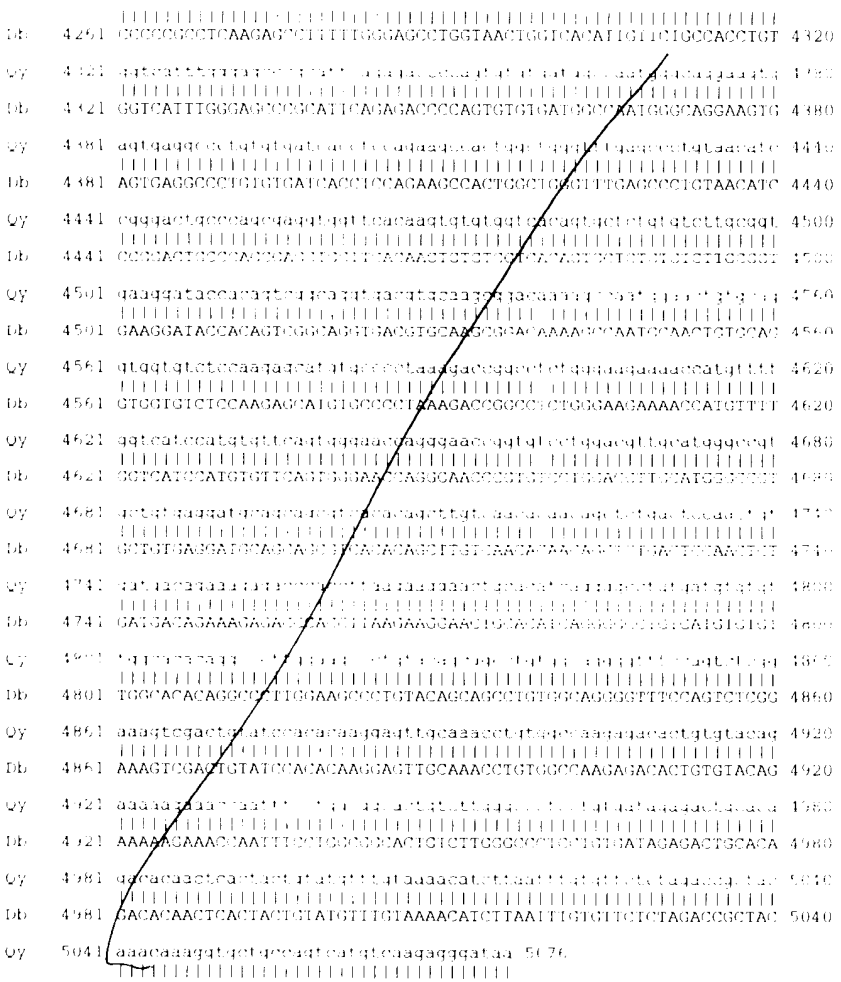
RESULT

Query Match 100.0%; Score 5076; DB 6; Length 5076;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 5076; Conservative 0; Mismatches 0; Indels 0; Gaps 0;





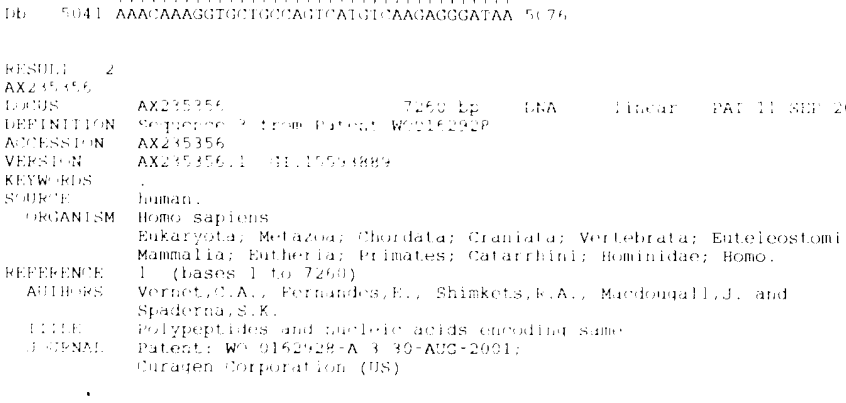




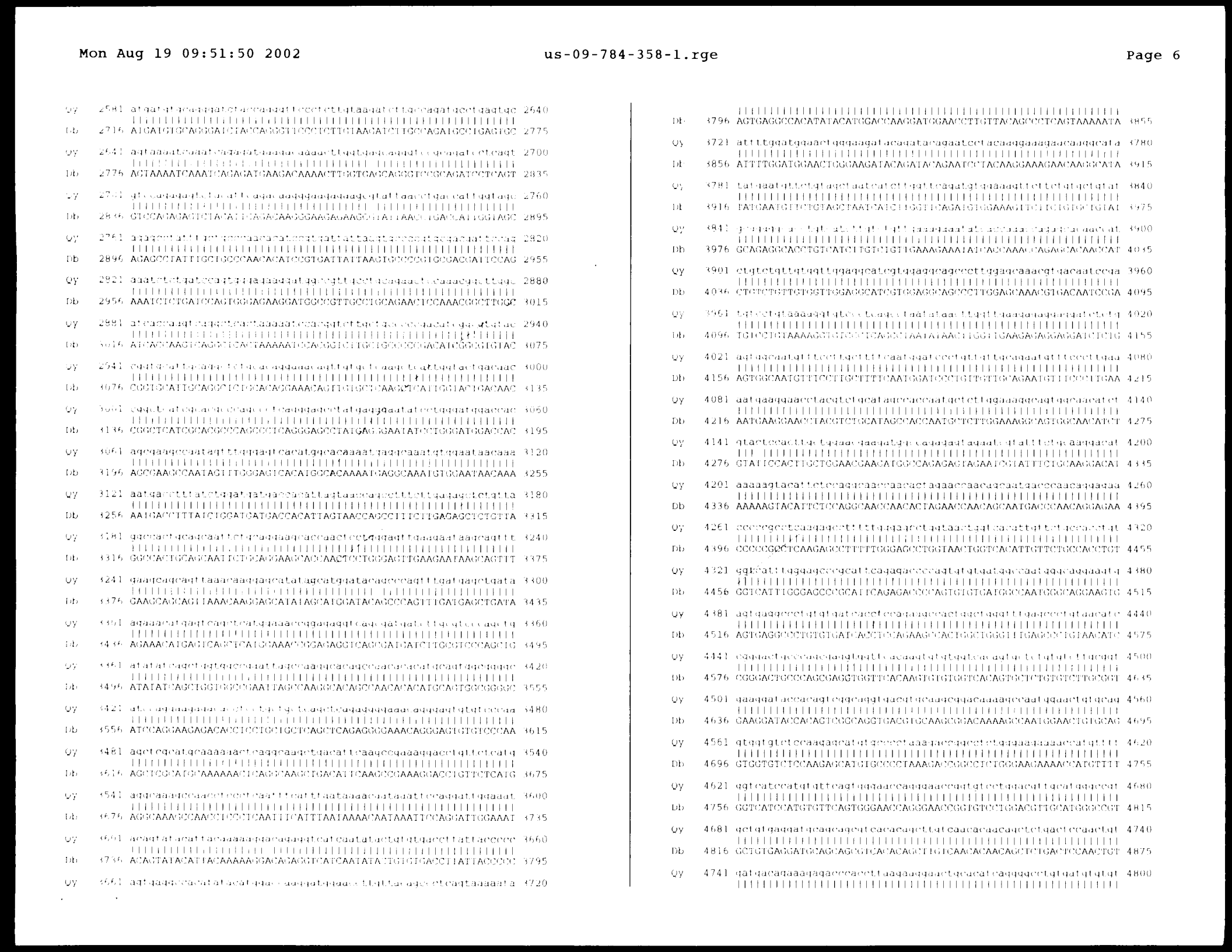
```
RESULT   2
AX235356
LOCUS       AX235356               7260 bp    DNA     linear   PAT 11-SEP-2001
DEFINITION  Sequence 2 from Patent WO0162928.
ACCESSION   AX235356
VERSION     AX235356.1  GI:15593889
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 7260)
  AUTHORS   Vernet,C.A., Fernandes,E., Shimkets,R.A., Macdougall,J. and
            Spaderna,S.K.
  TITLE     Polypeptides and nucleic acids encoding same
  JOURNAL   Patent: WO 0162928-A 3 30-AUG-2001;
            Curagen Corporation (US)
```

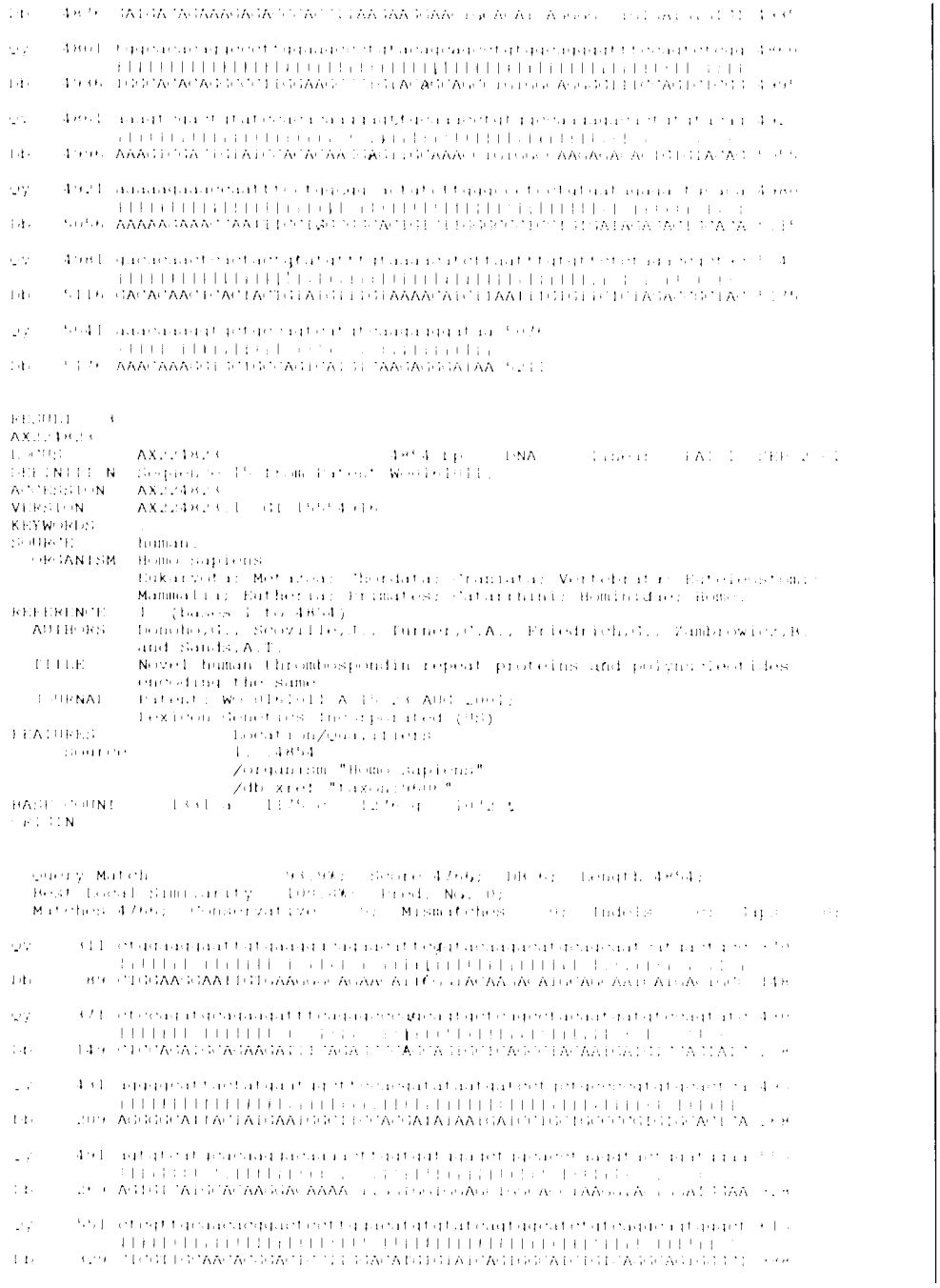
```
FEATURES             Location/Qualifiers
     source          1..7260
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             136..5211
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAC69812.1"
                     /db_xref="GI:15593890"
                     /translation="MASWTSPWWVLLGMVFMHSPLPQTTAEKSPGAYFLPEFALSPQG
                     SFLEDTTGFQFLTYRYDDQTSRNTESPEDKPCNWDAWCDWSDCSRTCGGCAGYSLRRC
                     LTGRNCEGQNIRYKTCSNHDCPPDAEDFRAQQCSAYNDVQYQGHYYEWLPRYNDPAAP
                     CAIKCHAQGQNLVVELAPKVLDGTRCNTDSLDMCISGICQAVGCDRQLGSNAKEDNCG
                     VCAGDGSTCPLVKGQFSAFSSDNFLRPMAGLVAEPITSTFYTAAEIGVVQFF
                     FYLPIPHLWFAIDTMISFFFIGVIDQFYLNGAILMSVDELPDLKFINVEIVKNPK
                     LKECSMDPCPSSDGFKEIMPYDHFQPILPRWFHNPWIACSVSCGGGIQRRSFVCVEESM
                     HCEITQVPFFWPCMYAPKPKVMQTCNIFVCPDPWIAMEWSQCTVTCGRGLRYRVVLCTNH
                     RGEHVGGCNPQLKPHIKFFCVTPTPCYKPEFKSHVEAFLPWLPQAPLEETRIATEEP
                     IEIPEPWSACSTTCGPGVQVREVKCRVLITFTQTETELPEEECEGPKLPTERPCTLEA
                     CDESPASKELDIPLPEDSETTYDWEYAGFTPCTATCVGGHQEAIAVCLHIQTQQTVNC
                     SLCDMVHRPAMSQACNTEPCPPRWHVGSWGPCSATCGVGIQTRDVYCLHPGETPAPP
                     EECRDEKPHALQACNQFDCPPGWHIPEWQCSRTCGGGTQNRPVTCRQLLTDGSFLNL
                     SDELCQGPKASSHKSCAFIECPPELAVGDWSPCSVSCGVGIQQRPFQVCQRLAAKGRRI
                     PLSEMMCRDLPGPPLVRSCQMPECSKIESEMFIKICEPGPQILSVQPVYIQTREEERI
                     NLIGSKEAYLSNISVIECDVSRPQFSLCQWEPIPCTCQNFEFLCTIPSPCLKIHGL
                     AAPDIGVYRCIAGSAQETVVLKLIGTDNRLIARPALREPMREYPQMDHSEANSLGVIW
                     HEMSQMWHDEGSVSAVLPRERQNLTLRAALSREHCKAGLNLWLLKNEVRLDGAAVRLSA
                     YSMDTAQFDELIKNMSQLMETGDVSDDLASPLIYQLVAELAKAQPTHMQWRGIQEEID
                     PAALWKREVSVQSSLAFNSIPIPSAAQKPVFVGMVRQGGLPSRGLNETINSKIGNIVYI
                     TKETEVINILTNLIPSEAIYTWLSLMPLQGSVFSWTDGSPVSFTNWKPGEPNNFRREQGIYE
                     QCVANELSAFVELDSVLPAEAFVTLSVEENITTEFHHNHLSVVVGGIVEAALGANVTLK
                     CPVEGVPQPNIIWLKKGVFLSGNVSLIFNCFLSQNVSLENEGIYVCTAENAIGKAVA
                     TSVFHLLERRWPSSRIVFTQCHKRYIQAINTRTNSNDPTGEPPPQEPEWEPQNWSHC
                     SATCGHLGARTQRFLCVMANSQFVGPAFCLGLPKPLASFEFCNTRFPFAMWFTSVWSQ
                     CSVSCGAGLRQRQVICERLEAGETVQVVLEPAFCAPESRPLQEESVIGHCVQWEGNR
                     CPGRCMGRAVRMQQRHTACQHNSSDGNCDERRFPTLRRRCTSGACDVCWHTGPWEPCT
                     AACGRCFQSRKVDCIHTPSCKPVAKPIDCVQKEKPISWRHCLGPSCDRDCTDTTHYCMF
                     VKHLNLCSILDRYKQRCCQSCQEG"
BASE COUNT      2022 a    1593 c    1811 g    1834 t
ORIGIN
```

Query Match 99.9%; Score 5663.6; DB 6; Length 7260;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 5072; Conservative 0; Mismatches 5; Indels 2; Gaps 0;





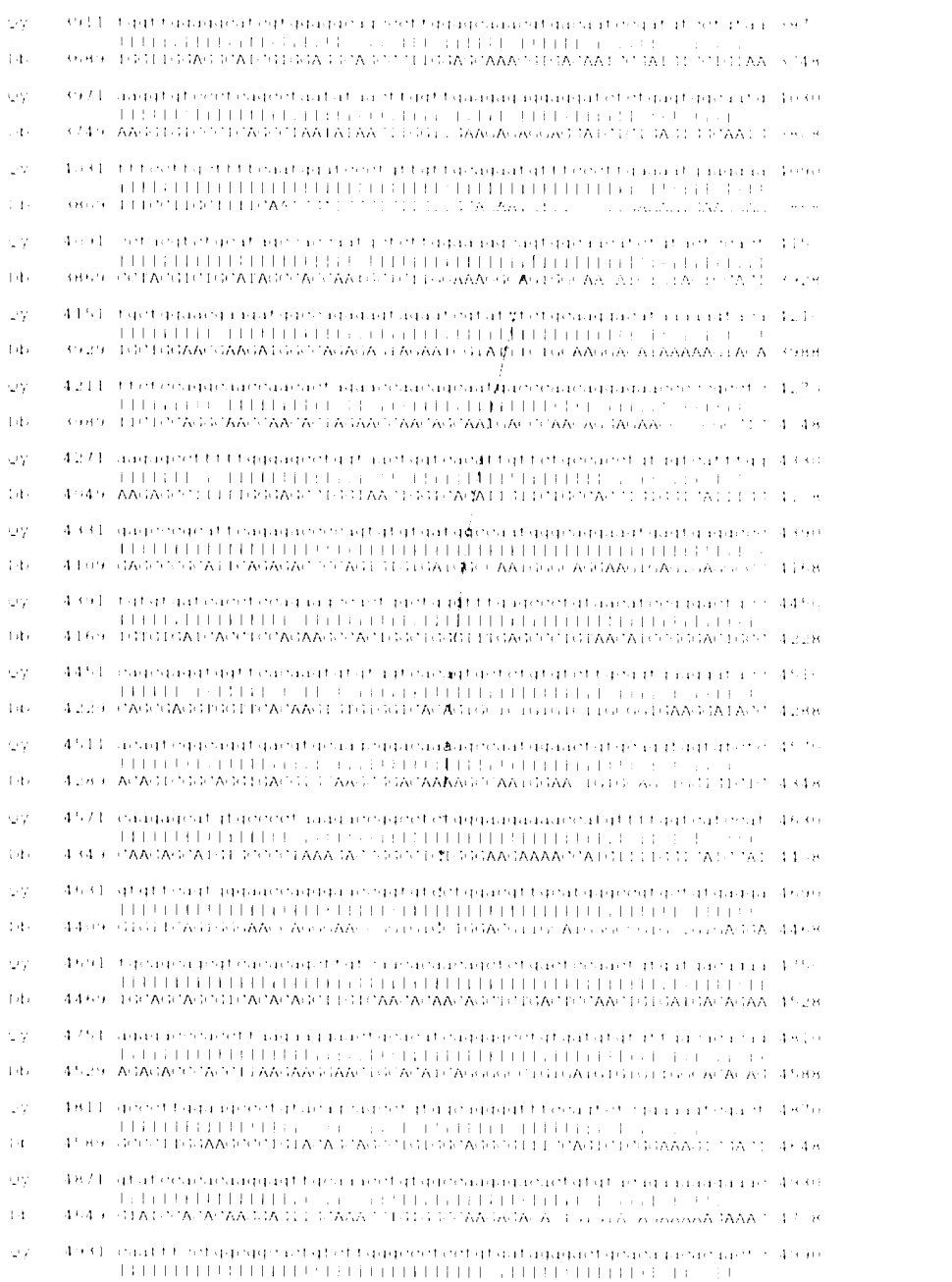




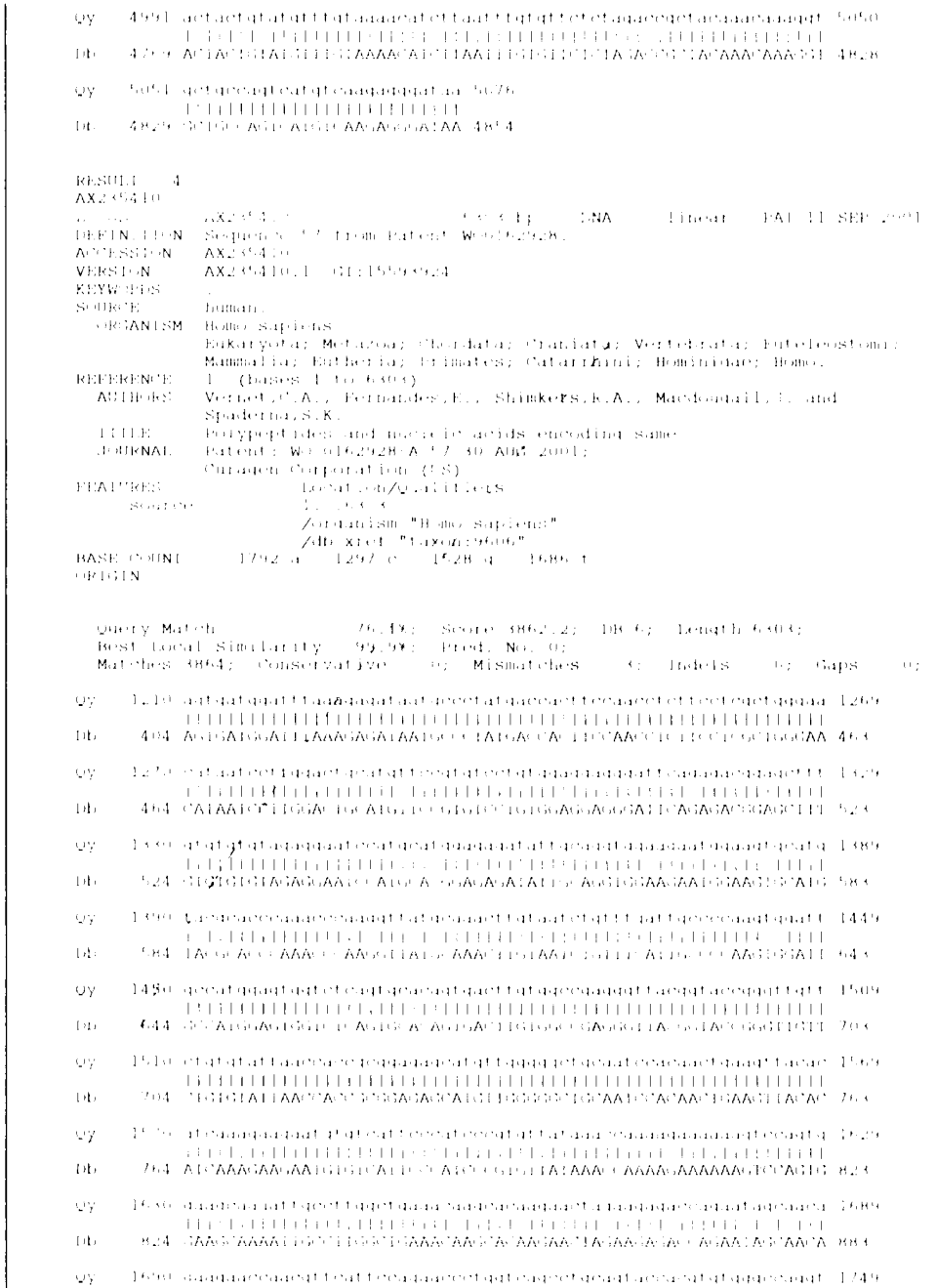
LOCUS 4
ACCESSION AX220828
DEFINITION Sequence 1 from Patent WO0164881.
ACCESSION AX220828
VERSION AX220828.1 GI:15554106
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 4854)
AUTHORS Donohoe,G., Scorzillo,T., Turner,C.A., Friedrich,G., Zambrowicz,B.
and Sands,A.T.
TITLE Novel human Thrombospondin repeat proteins and polynucleotides
encoding the same
JOURNAL Patent: WO 0164881-A 1 06-SEP-2001;
Lexicon Genetics Incorporated (US)
FEATURES Location/Qualifiers
source 1..4854
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT
ORIGIN

Query Match 93.9%; Score 4766; DB 6; Length 4854;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4766; Conservative 0; Mismatches 0; Indels 0; Gaps 0;





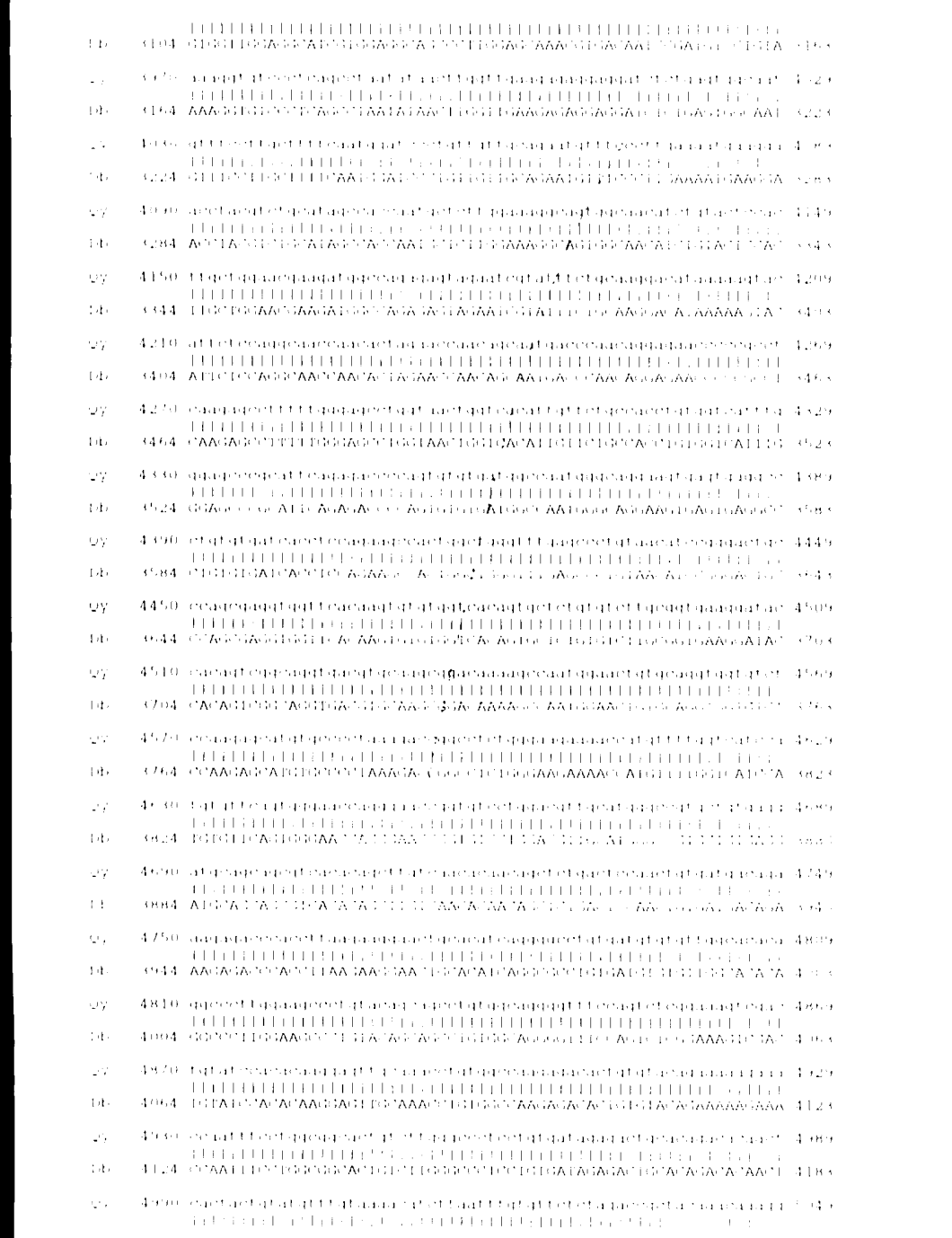
RESULT 4
AX295410
LOCUS AX295410 6303 bp DNA linear PAT 11-SEP-2001
DEFINITION Sequence 7 from Patent WO0162928.
ACCESSION AX295410
VERSION AX295410.1 GI:15599624
KEYWORDS .
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 6303)
AUTHORS Vernet,C.A., Fernandes,E., Shimkets,R.A., Macdonald,L. and
Spaderna,S.K.
TITLE Polypeptides and nucleic acids encoding same
JOURNAL Patent: WO 0162928-A 7 30-AUG-2001;
Curagen Corporation (US)
FEATURES Location/Qualifiers
source 1..6303
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 1792 a 1297 c 1528 g 1686 t
ORIGIN



Query Match 76.1%; Score 3862.2; DB 6; Length 6303;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 3864; Conservative 0; Mismatches 5; Indels 0; Gaps 0;







Db 4184 CACTACTGTATGTTTGTAAAACATTTAATTTGTGTTCTCTAGACCGCTACAAACAAAGG 4243

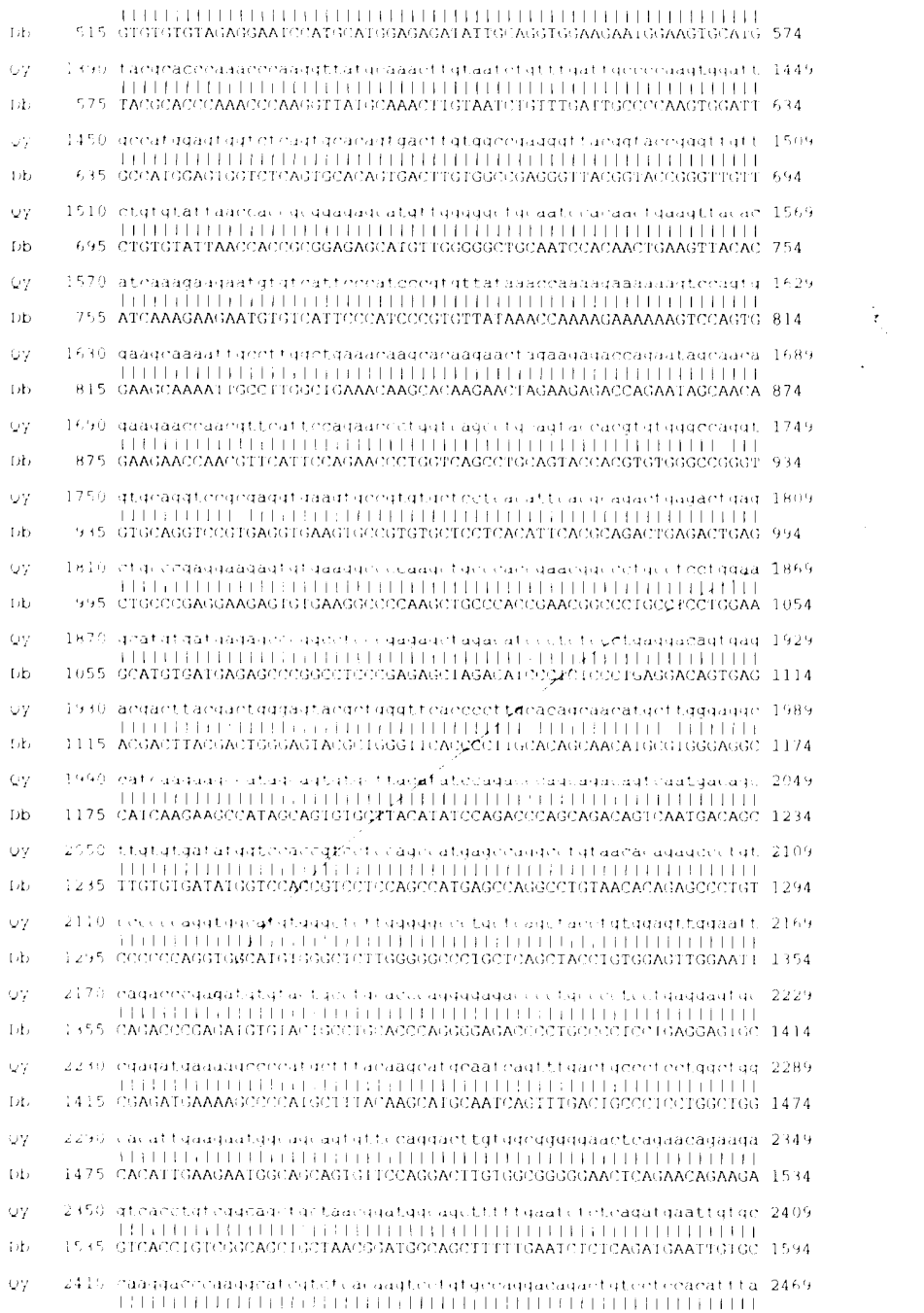
Qy 1210 aqtgatqtqatttaaaaqaqataatqcccctatqacactccaactcctcctcctqctqqqaa 1269

Db 395 ACTGGATGGATTTAAAGAGATAATCCCCTATGACCACTCCAACCTCTCCTCCTGGCGGGAA 454

Qy 1270 cataatcctqaactqcatqttcatqtcctatqaaaaaaatttcaaaaaqaaqqaqttt 1329

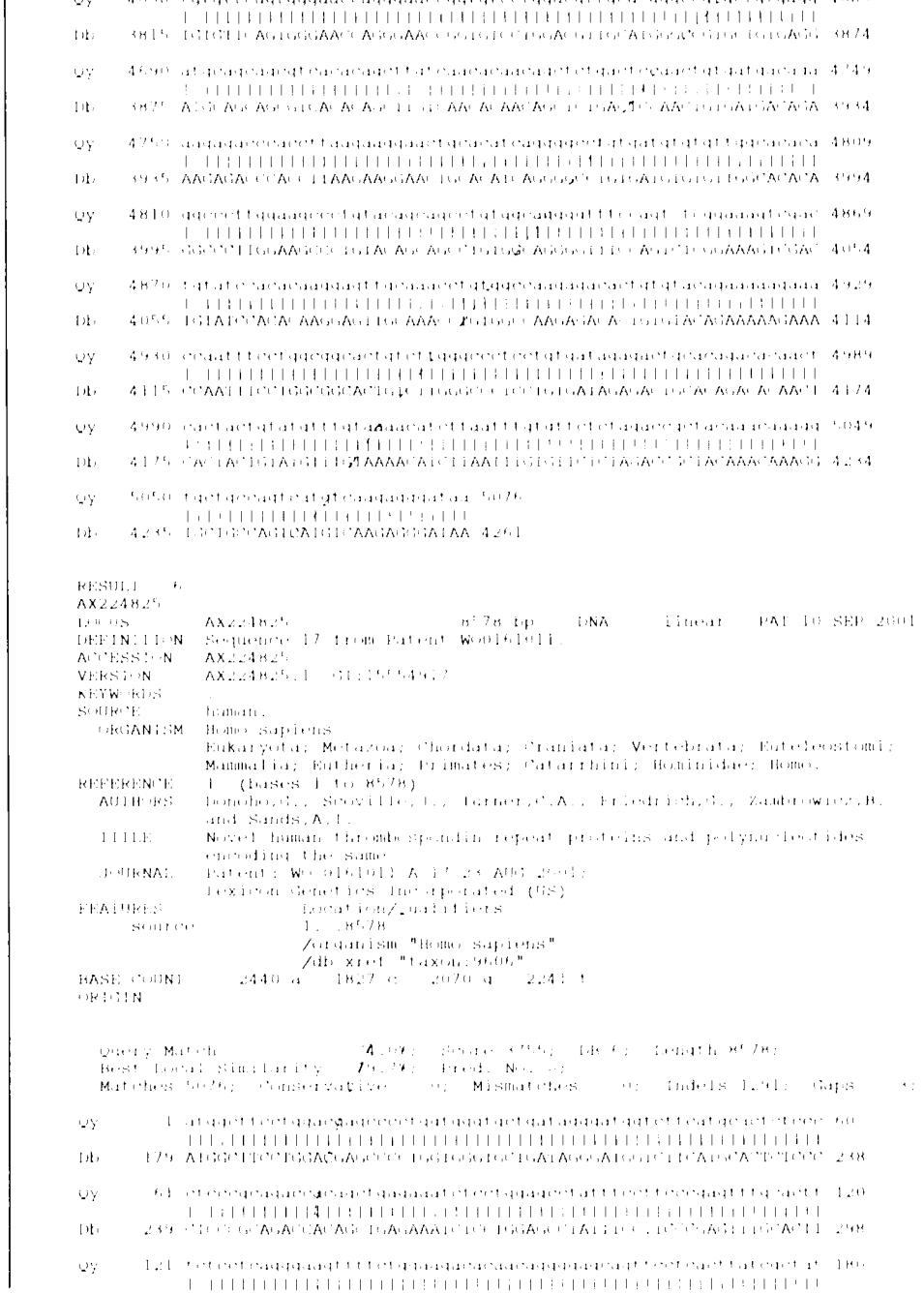
Db 455 CATAATCCTGGACTGCATGTTCATGTCCTATGAGGGGGTTTCAAAAAGAAGGAGTTT 514

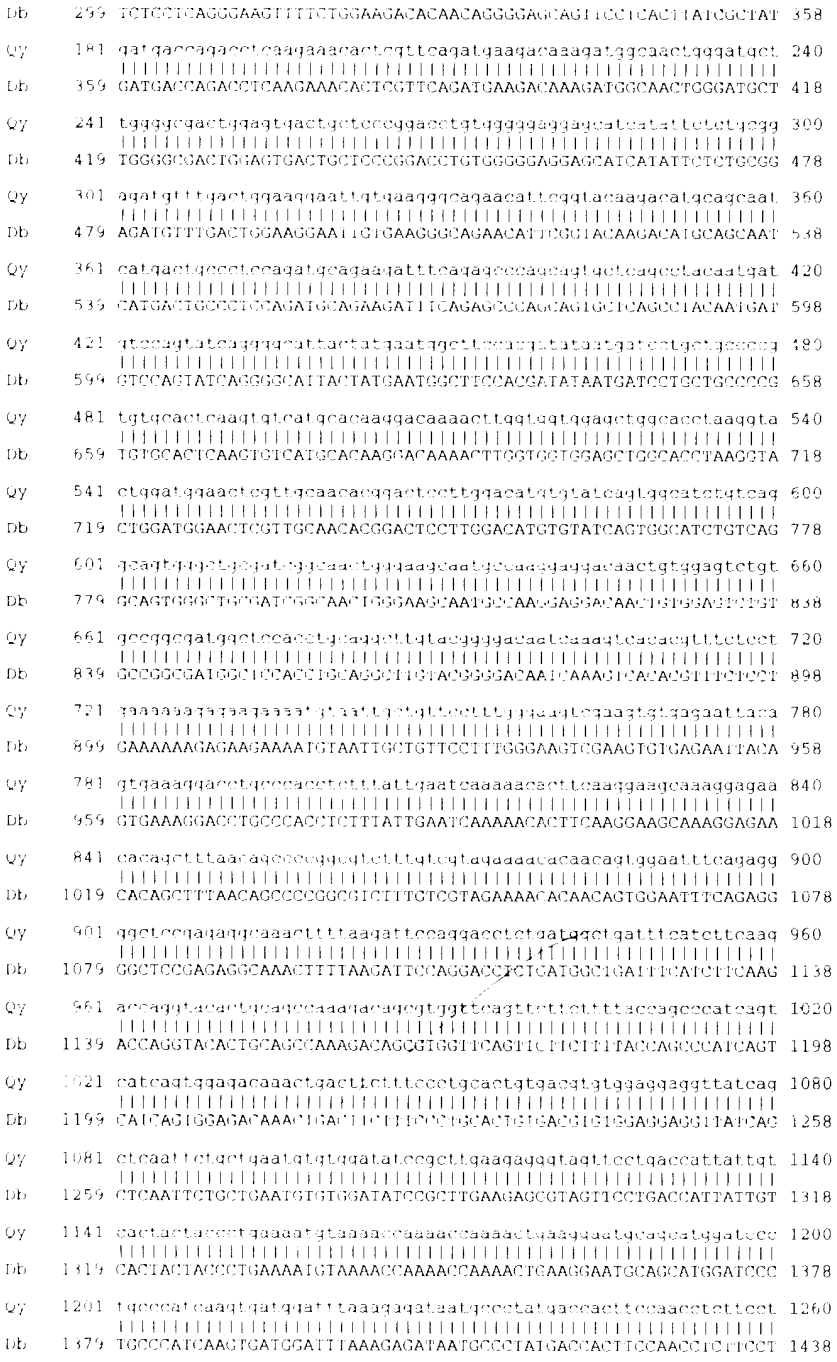
Qy 1330 tqtqtqtqaaaaaaatcaatqtqtqtqtqtqtqtqtqtqtqtqtqtqtqtqtqtqtq 1389

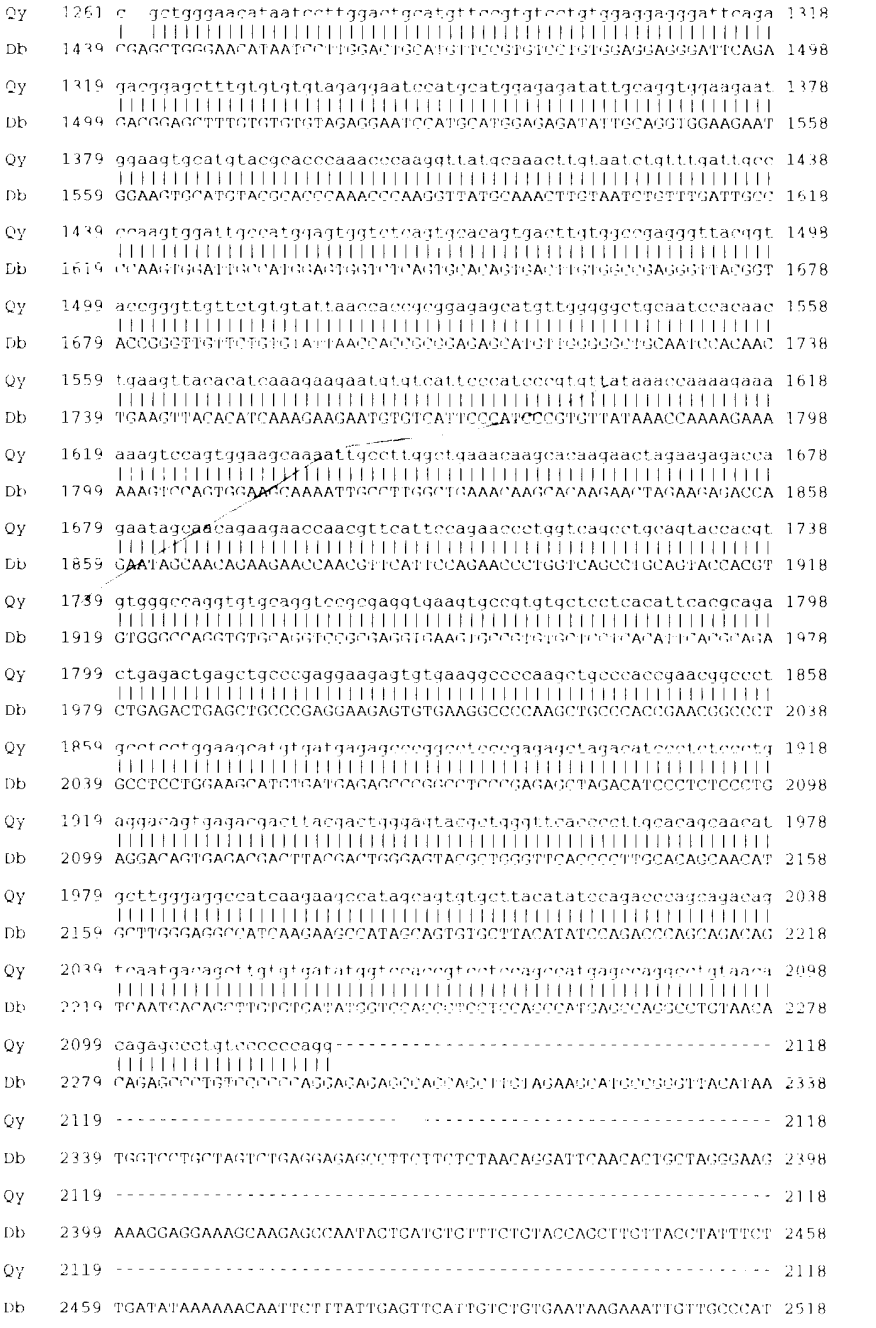


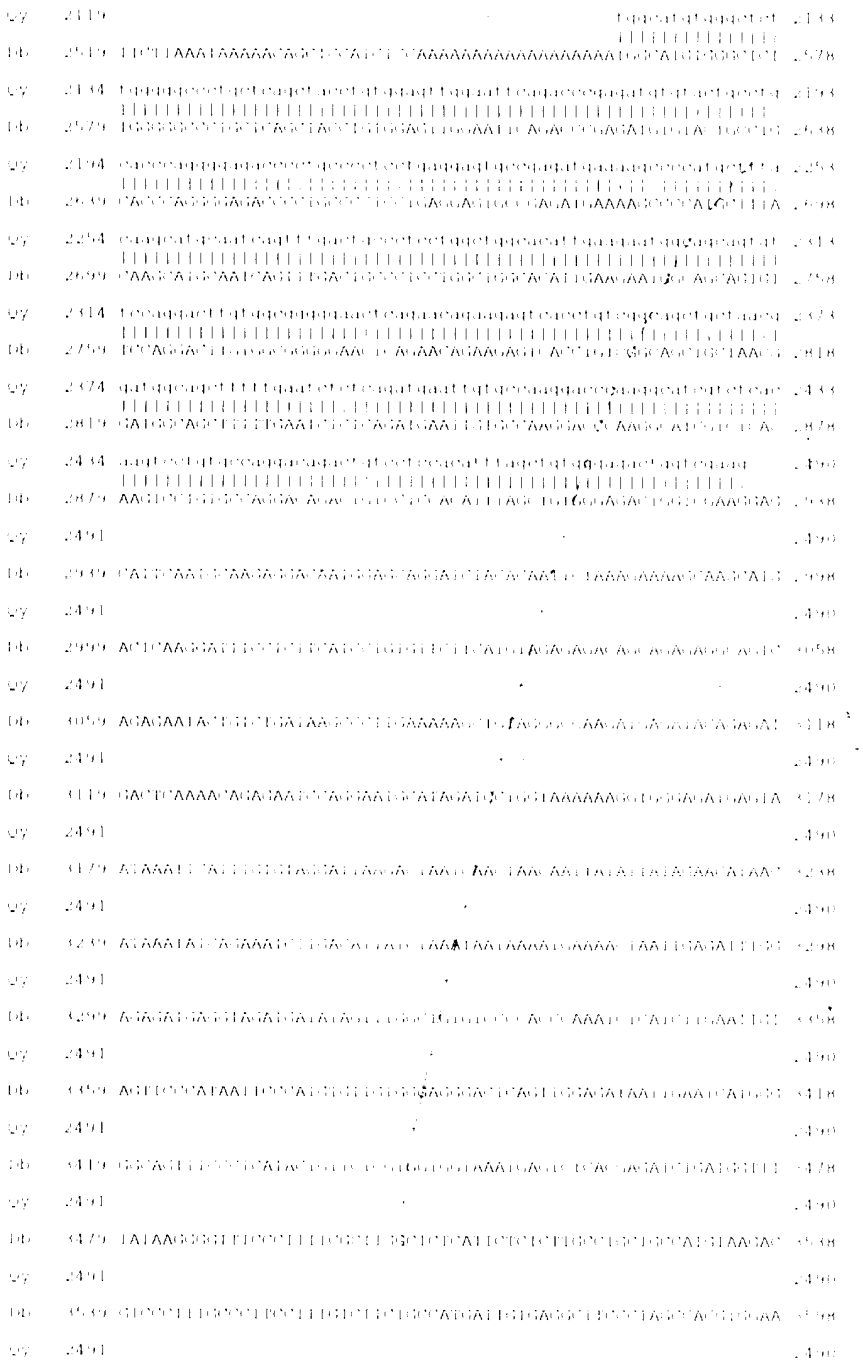


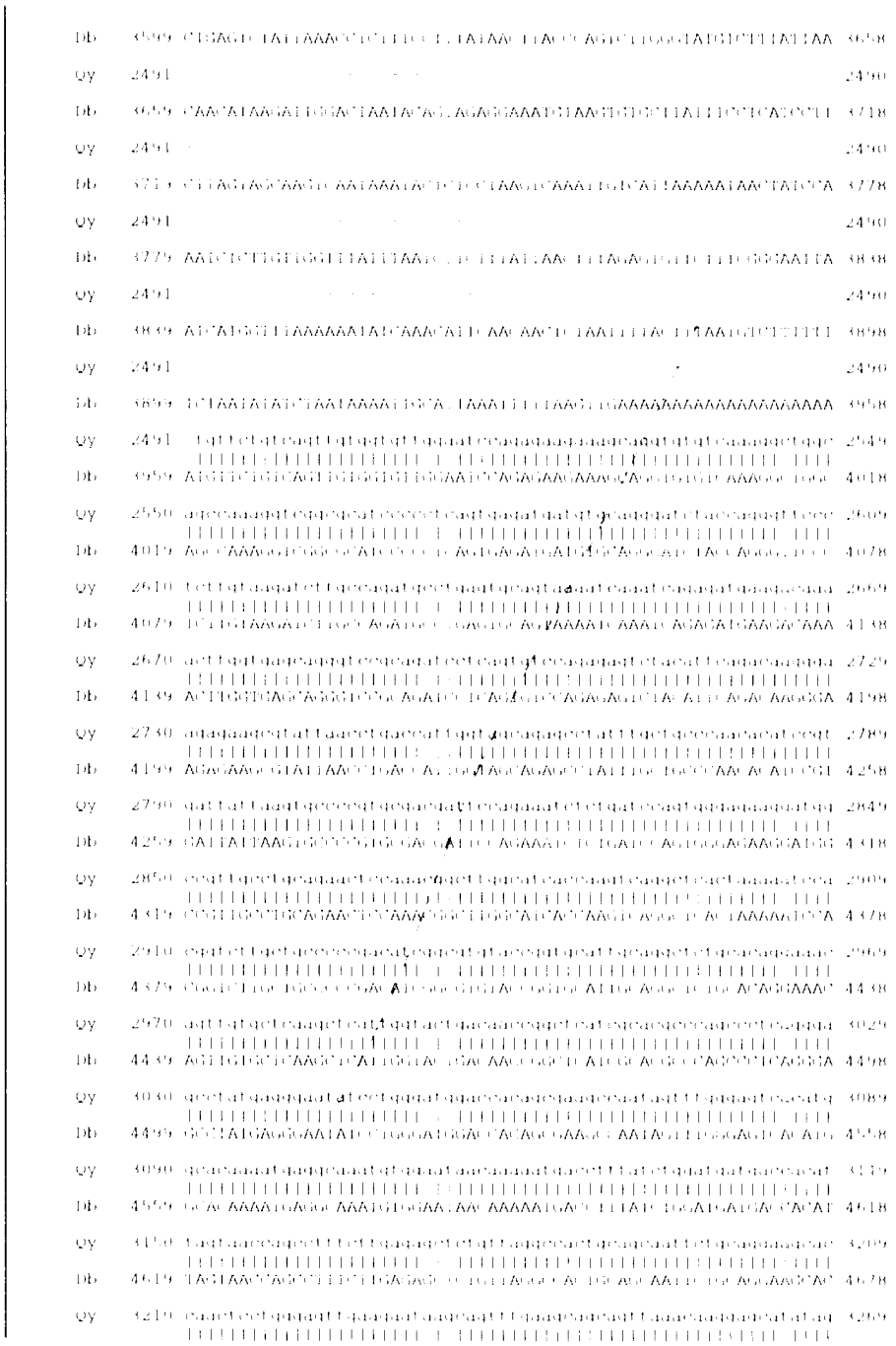
RESULT 6
AX224825
LOCUS AX224825 8578 bp DNA linear PAT 10-SEP-2001
DEFINITION Sequence 17 from Patent WO0161011.
ACCESSION AX224825
VERSION AX224825.1 GI:25954067
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 8578)
AUTHORS Donohoe,C., Scoville,L., Turner,C.A., Friedrich,G., Zaubrowicz,B.
and Sands,A.T.
TITLE Novel human thrombospondin repeat proteins and polynucleotides
encoding the same
JOURNAL Patent: WO 0161011 A 17-JAN-2001;
Lexicon Genetics Incorporated (US)
FEATURES Location/Qualifiers
source 1..8578
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 2440 a 1827 c 2070 g 2241 t
ORIGIN











```
Db    4679 CAACTCCTGGGAGTTGAAGAATAAGCAGTTTGAAGCAGCAGTTAAACAAGCAGCATATAG 4738

Qy    3270 catggatacagcccagtttgatgagctgataagaaacatgagtcagctcatggaaaccgg 3329
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4739 CATGGATACAGCCCAGTTTGATGAGCTGATAAGAAACATGAGTCAGCTCATGGAAACCGG 4798

Qy    3330 agaggtcagcgatgatctttgcgtcccagctgatatatcagctggtggccgaattagccaa 3389
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4799 AGAGGTCAGCGATGATCTTGCGTCCCAGCTGATATATCAGCTGGTGGCCGAATTAGCCAA 4858

Qy    3390 ggcacagccaacacacatgcagtggcgggcatccaggaagagacacctctgctgctca 3449
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4859 GGCACAGCCAACACACATGCAGTGGCGGGCATCCAGGAAGAGACACCTCCTGCTGCTCA 4918

Qy    3450 gctcagagggtgaaacagggagtgtgtcccaaagctcgcatgcaaaaaactcagggcaagct 3509
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4919 GCTCAGAGGGTGAAACAGGGAGTGTGTCCCAAAGCTCGCATGCAAAAAACTCAGGCAAGCT 4978

Qy    3510 gacattcaagccgaaaggacctgttctctatgaggcaaagcaaacctccctcaatttcatt 3569
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4979 GACATTCAAGCCGAAAGGACCTGTTCTCTATGAGGCAAAGCAAACCTCCCTCAATTTCATT 5038

Qy    3570 taataaaacaatgaaattccagtattgaaatacagtatacattacaaaaggacagaggt 3629
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5039 TAATAAAACAATAAATTCCAGGATTGGAAATACAGTATACATTACAAAAAGGACAGAGGT 5098

Qy    3630 catcaatatactgtgtgaccttattacccccagtgaggccacatatacatgaaccaaaga 3689
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5099 CATCAATATACTGTGTGACCTTATTACCCCCAGTGAGGCCACATATACATGAACCAAGGA 5158

Qy    3690 tgggacctgttacaagccctcagtaaaaatgatttggatggaaagatacagat 3749
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5159 TGGAACCTGTTACAGCCCTCAGTAAAAATGATTTGGATGGAACTGGGAAGATACAGAT 5218

Qy    3750 acagaatcctacaagtaaaagaacaaggcatatatgaatgttctgtagctaatcatcttgg 3809
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5219 ACAGAATCCTACAAGTAAAAGAACAAGGCATATATGAATGTTCTGTAGCTAATCATCTTGG 5278

Qy    3810 ttcagatgtggaaagttcttctgtgctgtatgcagaggcacctgtcatcttgtctgttga 3869
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5279 TTCAGATGTGGAAAGTTCTTCTGTGCTGTATGCAGAGGCACCTGTCATCTTGTCTGTTGA 5338

Qy    3870 aagaaatatcacccaaaccagagcacaaccatctgtctgttgtggttggaggcatctgtgga 3929
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5339 AAGAAATATCACCAAACCAGAGCACAACCATCTGTCTGTTGTGGTTGGAGGCATCTGTGGA 5398

Qy    3930 ggcagccctggaagcaaacgtgacaatccgatgtcctgtaaaaggtgtccctcagcctaa 3989
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5399 GGCAGCCCTTGGAGCAAACGTGACAATCCGATGTCCTGTAAAAGGTGTCCCTCAGCCTAA 5458

Qy    3990 tataacttggttgaagagaggaggatctctgagtggcaatgtttccttgcttttcaatgg 4049
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5459 TATAACTTGGTTGAAGAGAGGAGGATCTCTGAGTGGCAATGTTTCCTTGCTTTTCAATGG 5518

Qy    4050 atccctgttgttgcagaatgtttccttgaaaatgaaggaacctacgtctgcatagccac 4109
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5519 ATCCCTGTTGTTGCAGAATGTTTCCTTGAAAATGAAGGAACCTACGTCTGCATAGCCAC 5578

Qy    4110 caatgctcttggaaaggcaatgtgcaactatctgtactccacttgctggaacgaagatggcc 4169
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5579 CAATGCTCTTGGAAAGGCAGTGGCAACACTCGTACTCCACTTGCTGGAACGAAGATGGCC 5638

Qy    4170 agagagtagaatcgtatttctgcaaggacataaaaagtacattctccaggcaaccaacac 4229
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5639 AGAGAGTAGAATCGTATTTCTGCAAGGACATAAAAAGTACATTCTCCAGGCAACCAACAC 5698

Qy    4230 tagaaccaacagcaatgaccaacaggagaaccccgcctcaagagcctttttgggagcc 4289
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5699 TAGAACCAACAGCAATGACCCAACAGGAGAACCCCCGCCTCAAGAGCCTTTTTGGGAGCC 5758

Qy    4290 tggtaactggtcacattgttctgccacctgtggtcattgggagccccgcattcagagacc 4349
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5759 TGGTAACTGGTCACATTGTTCTGCCACCTGTGGTCATTGGGAGCCCCGCATTCAGAGACC 5818
```

```
Qy    4350 ccagtgtgtgatggccaatgggcaggaagtgagtgagggcctgtgtgatcacctccagaa 4409
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5819 CCAGTGTGTGATGGCCAATGGGCAGGAAGTGAGTGAGGGCCTGTGTGATCACCTCCAGAA 5878

Qy    4410 gccactggctgggtttgagccctgtaacatccgggactgcccagccgaggtggttcacaag 4469
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5879 GCCACTGGCTGGGTTTGAGCCCTGTAACATCCGGGACTGCCCAGCCGAGGTGGTTCACAAG 5938

Qy    4470 tgtgtggtcacagtgctctgtgtcttgcggtgaaggatgccacagtcggccaggtgaccttg 4529
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5939 TGTGTGGTCACAGTGCTCTGTGTCTTGCGGTGAAGGATGCCACAGTCGGCCAGGTGACCTTG 5998

Qy    4530 caagcggacaaaagccaatggaactgtgcaggtggtgtctccaaagagcatgtgccctaa 4589
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5999 CAAGCGGACAAAAGCCAATGGAACTGTGCAGGTGGTGTCTCCAAGAGCATGTGCCCTAA 6058

Qy    4590 agaccggcctctgggaagaaaaccatgtttggtcatccatgtgttcagtgggaaccagg 4649
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6059 AGACCGGCCTCTGGGAACAAAACCATGTTTGGTCATCCATGTGTTCAGTGGGAACCAGG 6118

Qy    4650 gaaccggtgtcctggacgttgcatgggcggtgctgtgaggatgcagcagcgtcacacagc 4709
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6119 GAACCGGTGTCCTGGACGTTGCATGGGCGGTGCTGTGAGGATGCAGCAGCGTCACACAGC 6178

Qy    4710 ttgtcaacacaacagctctgactccaactgtgatgacagaaagagacccacctaagaag 4769
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6179 TTGTCAACACAACAGCTCTGACTCCAACTGTGATGACAGAAAGAGACCCACCTTAAGAAG 6238

Qy    4770 gaactgcacatcagggggcctgtgatgtgtgttggcacacaggccctggaagccctgtac 4829
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6239 GAACTGCACATCAGGGGGCCTGTGATGTGTGTTGGCACACAGGCCCTGGAAGCCCTGTAC 6298

Qy    4830 agcagcctgtggcagggtttccagtctcggaaagtcgacttgtatccacacaaggagttg 4889
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6299 AGCAGCCTGTGGCAGGGTTTCCAGTCTCGGAAAGTCGACTTGTATCCACACAAGGAGTTG 6358

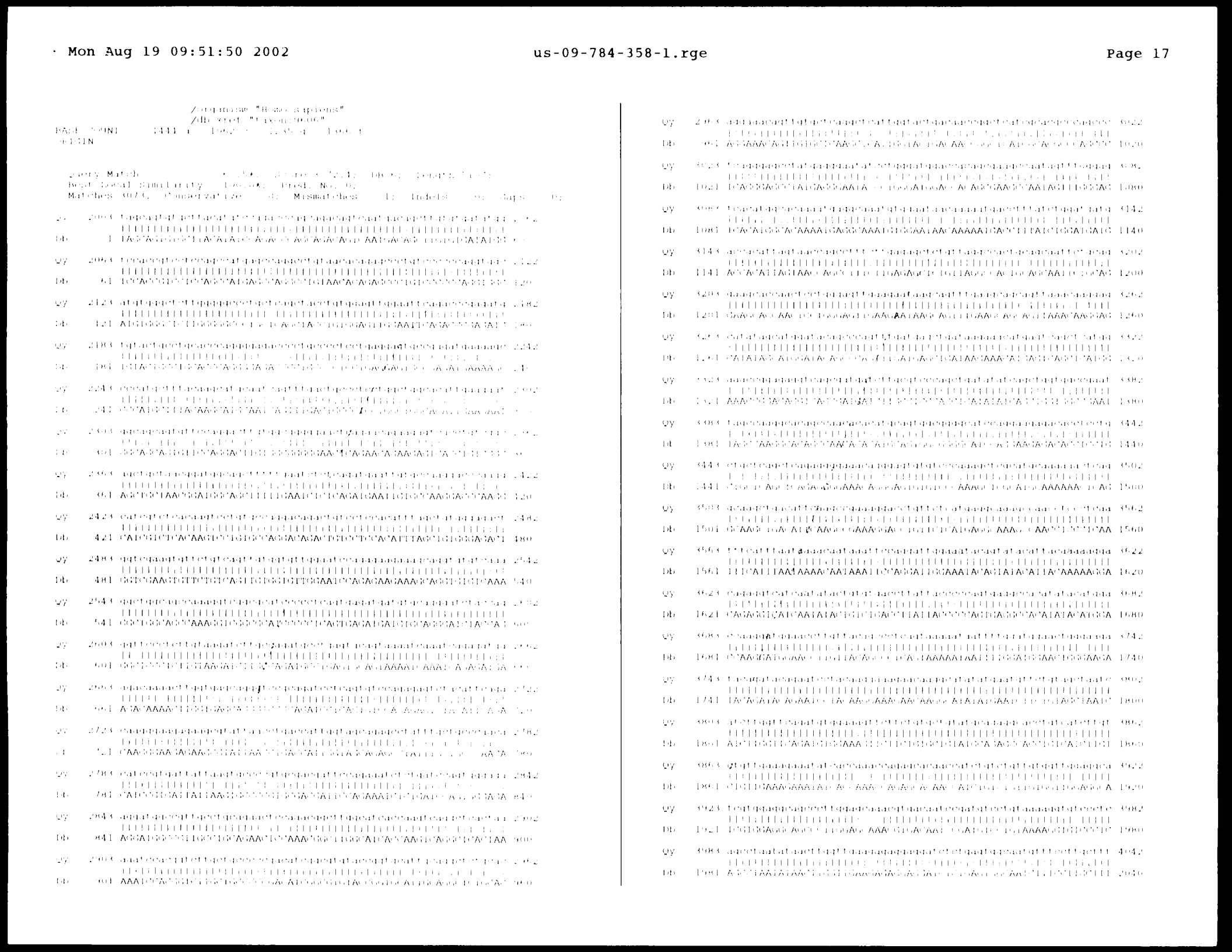
Qy    4890 caaacctgtggccaagagacactgtgtacagaaaaagaaaccaatttcctggcggcactg 4949
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6359 CAAACCTGTGGCCAAGAGACACTGTGTACAGAAAAAGAAACCAATTTCCTGGCGGCACTG 6418

Qy    4950 tcttgggccctcctgtgatagagactgcacagacacaactcactactgtatgtttgtaaa 5009
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6419 TCTTGGGCCCTCCTGTGATAGAGACTGCACAGACACAACTCACTACTGTATGTTTGTAAA 6478

Qy    5010 acatcttaatttgtgttctctagaccgctacaaacaaaggtgctgccagtcatgtcaaga 5069
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6479 ACATCTTAATTTGTGTTCTCTAGACCGCTACAAACAAAGGTGCTGCCAGTCATGTCAAGA 6538

Qy    5070 qggataa 5076
           |||||||
Db    6539 GGGATAA 6545
```

```
RESULT   7
AX235377
LOCUS       AX235377               5107 bp    DNA     linear   PAT 11-SEP-2001
DEFINITION  Sequence 24 from Patent WO0162928.
ACCESSION   AX235377
VERSION     AX235377.1  GI:15593909
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 5107)
  AUTHORS   Vernet,C.A., Fernandes,F., Shimkets,R.A., Macdougall,J. and
            Spaderna,S.K.
  TITLE     Polypeptides and nucleic acids encoding same
  JOURNAL   Patent: WO 0162928 A 24 30 AUG-2001;
            Curagen Corporation (US)
FEATURES             Location/Qualifiers
     source          1..5107
```



```
Qy    4041 tcaatqqatccctqttattqcagaatqtttcccttqaaaatqaatqaacctacqtctqca 4102
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2041 TCAATGGATCCCTGTTGTTGCAGAATGTTTCCCTTGAAAATGAAGGAACCTACGTCTGCA 2100

Qy    4103 tagccacccaatqctcttqtaaaaqtttaqtaaaaatttattcttqqtqtaaqtaa 4162
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2101 TAGCCACCAATGCTCTTGGAAAGGCAGTGGCAACATCTGTACTCCACTTGCTGGAAGGAA 2160

Qy    4163 qatqqccaqaqaqtaqaaatcqtatttctqcaaqqacatacaaaaaqtacattctcaqqcaa 4222
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2161 GATGGTCAGAGAGTAGAATCGTATTTCTGCAAGGACATACAAAAAGTACATTCTCCAGGCAA 2220

Qy    4223 ccaaacactaqaaccaacaqcaatqaccccaacaqqaqaaccccqcctcaaqaqccttttt 4282
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2221 CCAACACTAGAACCAACAGCAATGACCCCAACAGGAGAACCCCCGCCTCAAGAGCCTTTTT 2280

Qy    4283 qqaaaccbatcaactatcaaatattttqtqccacctqtqqtcattttqqaaqccqqcattc 4342
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2281 GGGAGCCTGGTAACTGGTCACATTGTTCTGCCACCTGTGGTCATTTGGGAGCCCGCATTC 2340

Qy    4343 aqaqaccccaqtqtqtqatqaqcaaqaqqcaqqaaqtqaatqaaaqccctqtqtqtqatcaca 4402
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2341 AGAGACCCCAGTGTGTGATGGCCAATGGGCAGGAAGTGAGTGAGGCCCTGTGTGATCACC 2400

Qy    4403 tccaqaaqccactqqqtqqtttqaqcccqtqaacatccqqqactgccaaccaqaqqtqqt 4462
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2401 TCCAGAAGCCACTGGGTGGTTTGAGCCCTGTAACATCCGGGACTGCCCAGCAGAGGTGGT 2460

Qy    4463 tcaaaqtqtqtqqtcacaqtqcctqtatctqtqqtqaaqqatqccacaqtcqqcaqq 4522
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2461 TCACAAGTGTGTGGTCACAGTGCCTCTGTGTCTTGCGGTGAAGGATACCACAGTCGGCAGG 2520

Qy    4523 tqaccqtqccaaqcqqaccaaaaqccaatqqaactqtqccaqtqqtqtqtccaaaaqcatqtq 4582
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2521 TGACGTGCAAGCGGACCAAAAGCCAATGGAACTGTGCCAGTGGTGTCTCCAAGAGCATGTG 2580

Qy    4583 cccctaaaqacqqqcctctqqqaaqaaaaccatqttttqtqtcatcatqtqttcaqtqqq 4642
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2581 CCCCTAAAGACCGGCCTCTGGGAAGAAAACCATGTTTTGGTCATCCATGTGTTCAGTGGG 2640

Qy    4643 aacccatqqaaccqqtqtcctqqacqttqcatqqqccqtqctqtqaqqatqcaqcaqcqtc 4702
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2641 AACCCAGGGAACCGGTGTCCTGGACGTTGCATGGGCCGTGCTGTGAGGATGCAGCAGCGTC 2700

Qy    4703 acacaqcttqtcaaccacaaccaqctctqactccaactqtqatqacacqaaaqaqacccaqct 4762
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2701 ACACAGCTTGTCAACCACAACCAGCTCTGACTCCAACTGTGATGACACGAAAGAGACCCAGCT 2760

Qy    4763 taaqaaqqaaactqcacatcaqqqctqtqatqtqtqttqccacaqqqqcctttqqaaqc 4822
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2761 TAAGAAGGAACTGCACATCAGGGCCTGTGATGTGTGTTGCCACACAGGCCCTTGGAAGC 2820

Qy    4823 cctqtacaqcaqcctctqqqcaqqqttttccaqttctcqqaaaqtcqactqtatccacqcaa 4882
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2821 CCTGTACAGCAGCCTGTGGCAGGGGTTTCCAGTCTCGGAAAGTCGACTGTATCCACACAA 2880

Qy    4883 qqaqttqccaaacctqtqqccaaqaqacacactqtqtacqqaaaaaqqaaccaatttcctqqc 4942
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2881 GGAGTTGCAAACCTGTGGCCAAGAGACACTGTGTACGAAAAAGAAACCAATTTCCTGGC 2940

Qy    4943 qqcactqtctqqqcccctcctqtqtatqqaqactqccaqaqcaaqtccactatqtatqt 5002
           .||||||||.|||.|.||||.|||||.|||||||||||.|||||||.|||||.||||||||||
Db    2941 GGCACTGTCTTGGGCCCTCCTGTGATAGAGACTGCACAGACACAACTCACTACTGTATGT 3000

Qy    5003 ttqtaaaaacatcttaatttatqttctctaqaccqctacaaacaaaqqtqctqccaqtcat 5062
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3001 TTGTAAAACATCTTAATTTGTGTTCTCTAGACCGCTACAAACAAAGGTGCTGCCAGTCAT 3060

Qy    5063 qtccaaqaqqqataa 5078
           ||||||||||||||||
Db    3061 GTCAAGAGGGATAA 3074
```

```
RESULT   8
AX235379
LOCUS       AX235379                5107 bp    DNA     linear   PAT 11-SEP-2001
DEFINITION  Sequence 26 from Patent WO0162928.
ACCESSION   AX235379
VERSION     AX235379.1  GI:15593910
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 5107)
  AUTHORS   Vernet,C.A., Fernandes,E., Shimkets,R.A., Macdougall,J. and
            Spaderna,S.K.
  TITLE     Polypeptides and nucleic acids encoding same
  JOURNAL   Patent: WO 0162928-A 26 30 AUG-2001;
            Curagen Corporation (US)
FEATURES             Location/Qualifiers
     source          1..5107
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT     1444 a   1062 c   1235 g   1366 t
ORIGIN
```

Query Match 60.5%; Score 3072.4; DB 6; Length 5107;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3073; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy    2003 tqcaqtqtqcttaatatcaqaccaaqcaqacaqtcaatqacaqcttqtqtqatatqq 2062
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 TAGCAGTGTGCTTACATATCCAGACCCAGCAGACAGTCAATGACAGCTTGTGTGATATGG 60

Qy    2063 tccaccqtqcctccaqccatqaqccaqqccctqtaacacaqaqccctqtcccccaqqtqqc 2122
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 TCCACCGTGCCTCCAGCCATGAGCCAGGCCCTGTAACACAGAGCCCTGTCCCCCAGGTGGC 120

Qy    2123 atqtqqqctcttqqqqqcccqtctcaqctcactqtqaqttqaqatttcaqaccqaqatq 2182
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 ATGTGGGCTCTTGGGGGCCCTGCTCAGCTACCTGTGGAGTTGGAATTCAGACCCGAGATG 180

Qy    2183 tqtactqcctqtqaccccaqqqaqaqaccccqtqcccctcctqaqqaqtqccqqaqatqaaaaqc 2242
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 TGTACTGCCTGCACCCAGGGGAGACCCCTGCCCCTCCTGAGGAGTGCCGAGATGAAAAGC 240

Qy    2243 ccctatcttacaaaqatqcaatcaqttqactqcctcctqqctqccacattqaaqaat 2302
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 CCCATGCTTTACAAGCATGCAATCAGTTTGACTGCCTCCTGGCTGCCACATTGAAGAAT 300

Qy    2303 qqcaqcaqtqtttcaqqactttqtqaqaaaaaatcaqqaqcaaaaqtqaccatqtqq 2362
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 GGCAGCAGTGTTCCAGGACTTTGTGGCGGGGAACTCAGAACAGAAGAGTCACCTGTCGG 360

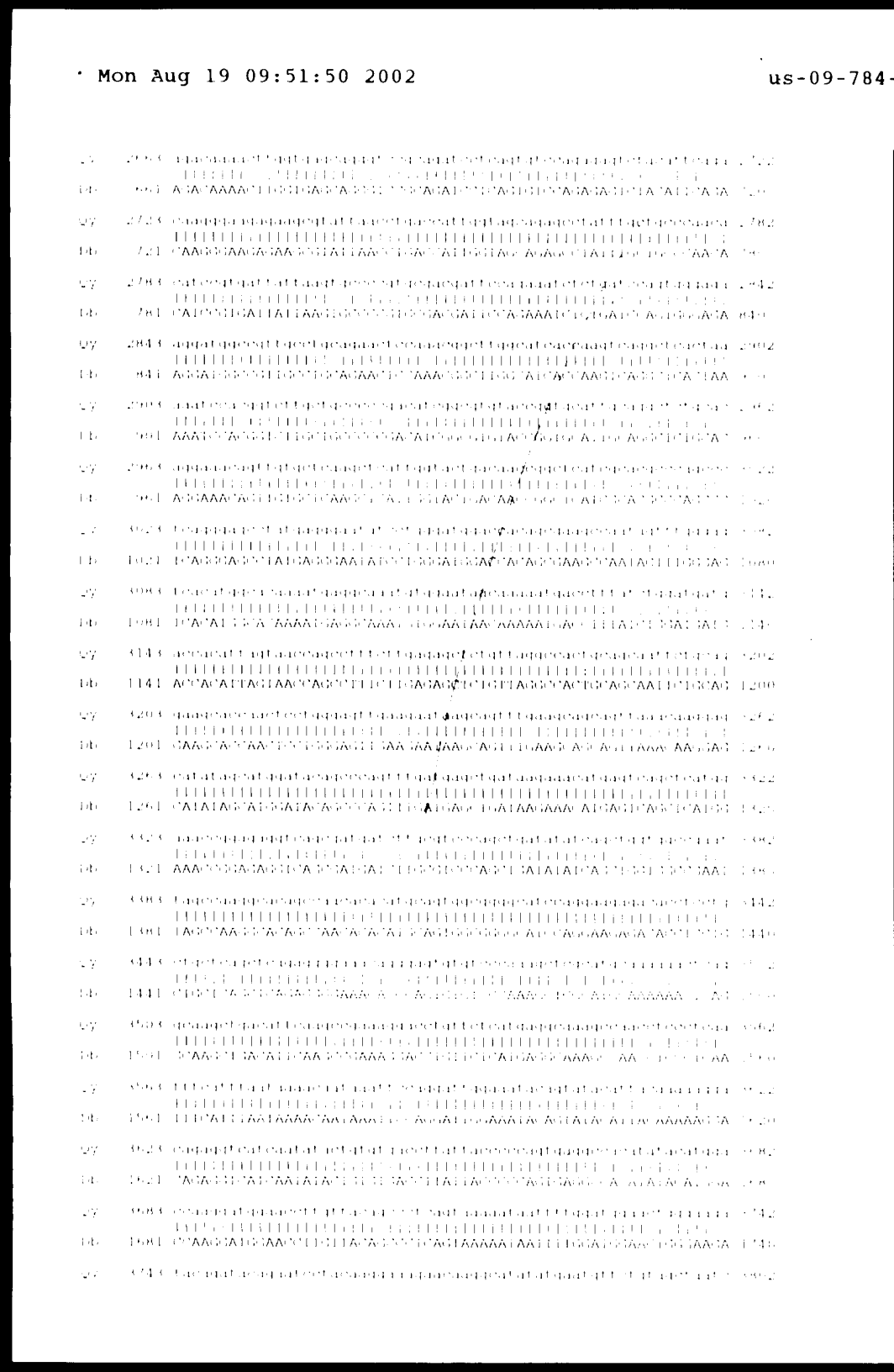
Qy    2363 aqctqctaaccqqatqtqaqctttttqaatctctcaqatqaattqtqccaqqqccaqq 2422
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 AGCTGCTAACGGATGCAGCTTTTTGAATCTCTCAGATGAATTGTGCCAAGGACCCAAGG 420

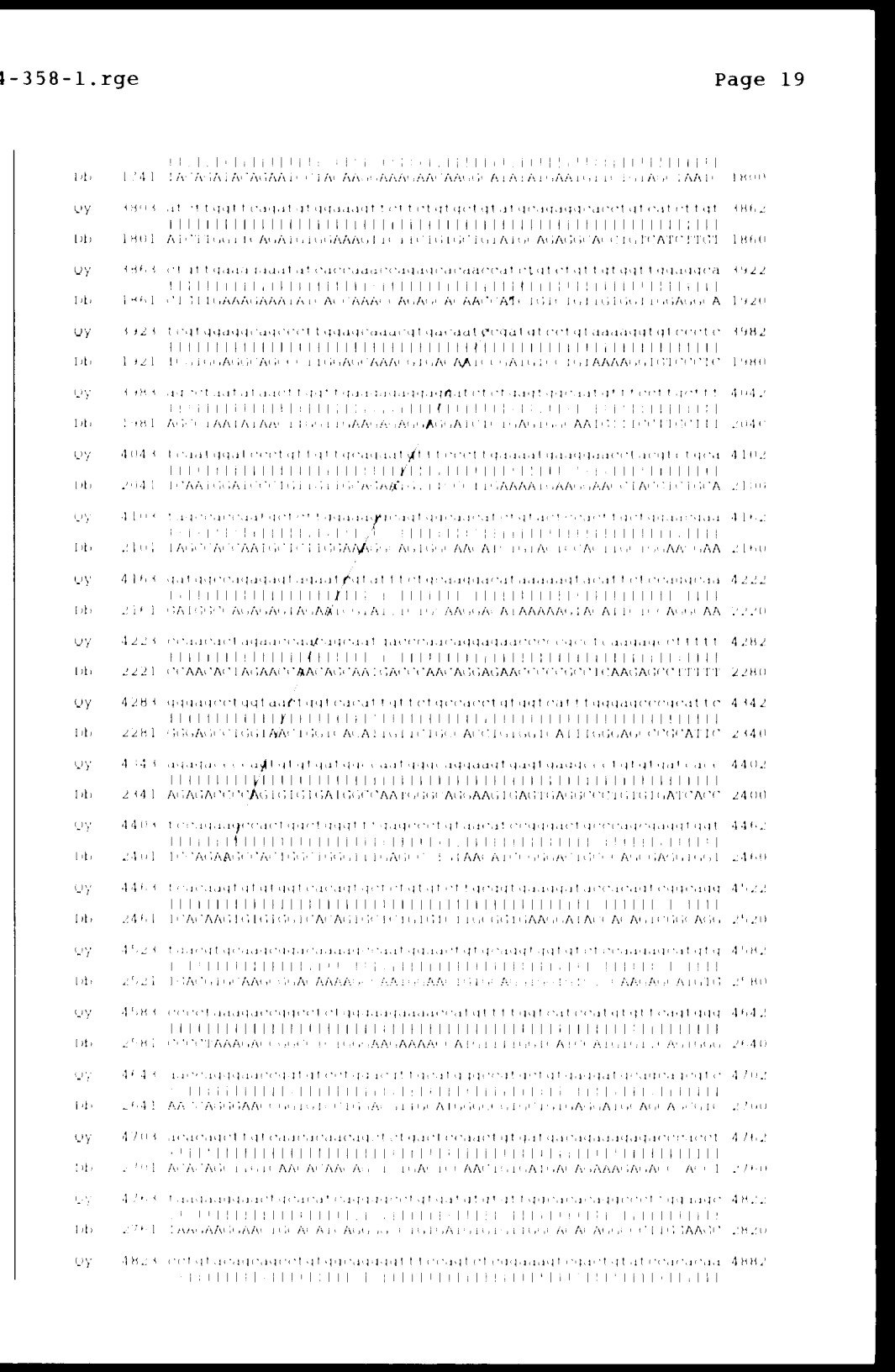
Qy    2423 catcqtctcacaaqtcctqtqccaqqacaqacqtqtcctccacatttaqctqtqqqaqact 2482
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 CATCGTCTCACAAGTCCTGTGCCAGGACAGACTGTCCTCCACATTTAGCTGTGGGAGACT 480

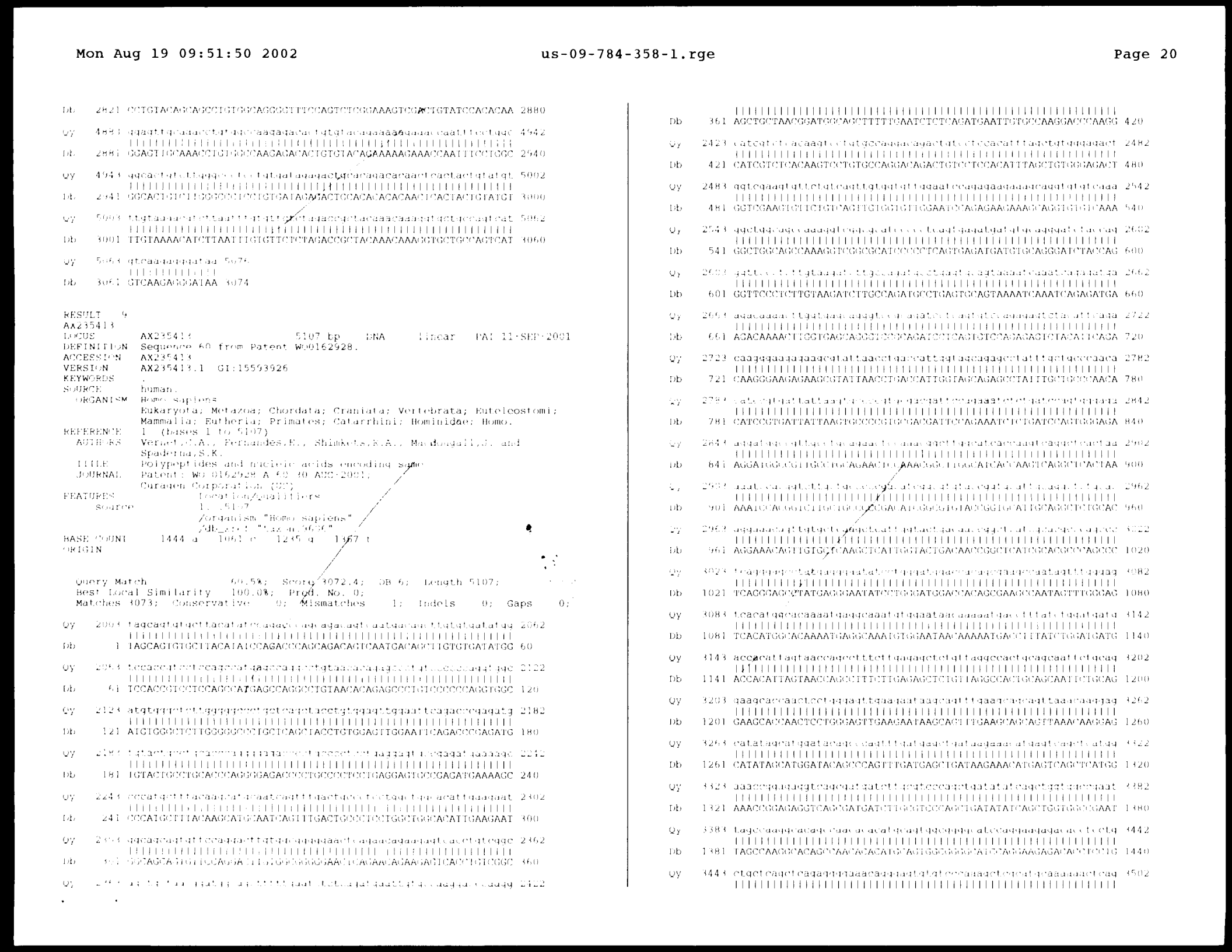
Qy    2483 qqtccaaqtqtttctqtcaqttqtqqtattqcaaatccqqaqaaqqaaqccaqqtqtqtccaaa 2542
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 GGTCGAAGTGTTCTGTCAGTTGTGGTGTTGGAATCCAGAGAAGAAAGCAGGTGTGTCAAA 540

Qy    2543 qqctqqcaqccaaaaqtcqqcqcatcccctcaqtqaqatqatqtqcaqqqatctaccaq 2602
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 GGCTGGCAGCCAAAGGTCGGCGCATCCCCCTCAGTGAGATGATGTGCAGGGATCTACCAG 600

Qy    2603 qttccctctttqtaaqatctqccaqatqccqaqtqcaqtaaaatcaaatcaqqqatqa 2662
           || ||||||||||:||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 GGCTCCCTCTTGTAAGATCTTGCCAGATGCCGAGTGCAGTAAAATCAAATCAGGAGATGA 660
```



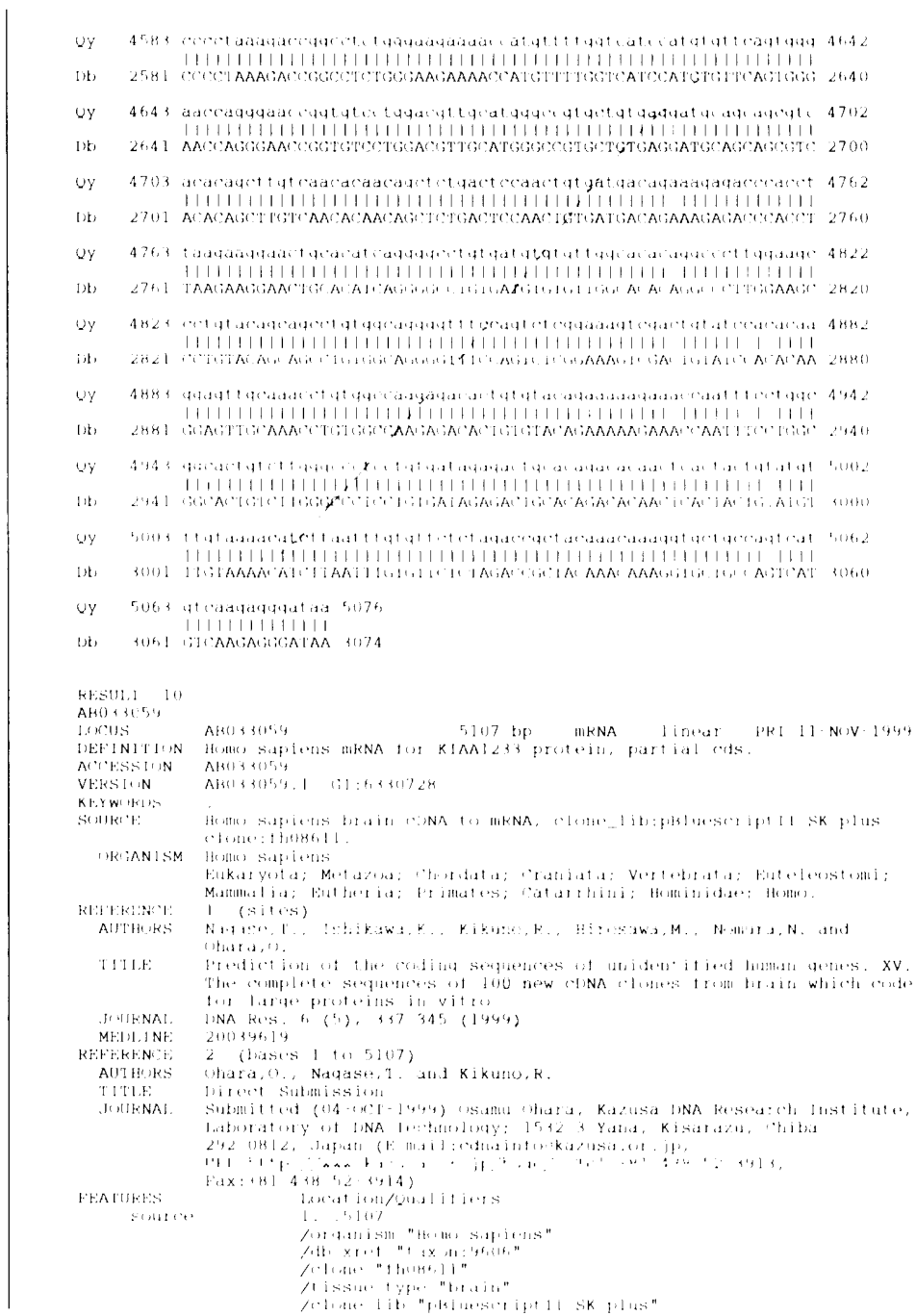




RESULT 9
AX235413
LOCUS AX235413 5107 bp DNA linear PAT 11-SEP-2001
DEFINITION Sequence 60 from Patent WO0162928.
ACCESSION AX235413
VERSION AX235413.1 GI:15593926
KEYWORDS .
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 5107)
AUTHORS Vernet,C.A., Fernandes,E., Shimkets,R.A., Macdougall,J. and
Spaderna,S.K.
TITLE Polypeptides and nucleic acids encoding same
JOURNAL Patent: WO 0162928 A 60 30 AUG-2001;
Curagen Corporation (US)
FEATURES Location/Qualifiers
source 1..5107
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 1444 a 1061 c 1245 g 1357 t
ORIGIN

Query Match 60.5%; Score 3072.4; DB 6; Length 5107;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3073; Conservative 0; Mismatches 1; Indels 0; Gaps 0;





```
RESULT  10
AB033059
LOCUS       AB033059                5107 bp    mRNA    linear   PRI 11-NOV-1999
DEFINITION  Homo sapiens mRNA for KIAA1233 protein, partial cds.
ACCESSION   AB033059
VERSION     AB033059.1  GI:6330728
KEYWORDS    .
SOURCE      Homo sapiens brain cDNA to mRNA, clone_lib:pBluescriptII SK plus
            clone:fh08611.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Nagase,T., Ishikawa,K., Kikuno,R., Hirosawa,M., Nomura,N. and
            Ohara,O.
  TITLE     Prediction of the coding sequences of unidentified human genes. XV.
            The complete sequences of 100 new cDNA clones from brain which code
            for large proteins in vitro
  JOURNAL   DNA Res. 6 (5), 337-345 (1999)
  MEDLINE   20089619
REFERENCE   2  (bases 1 to 5107)
  AUTHORS   Ohara,O., Nagase,T. and Kikuno,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-OCT-1999) Osamu Ohara, Kazusa DNA Research Institute,
            Laboratory of DNA Technology; 1532-3 Yana, Kisarazu, Chiba
            292-0812, Japan (E-mail:dnainfo@kazusa.or.jp,
            [illegible]
            Fax:081-438-52-3914)
FEATURES             location/Qualifiers
     source          1..5107
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="fh08611"
                     /tissue_type="brain"
                     /clone_lib="pBluescriptII SK plus"
```

```
     gene            1..3074
                     /gene "KIAA1233"
     CDS             1..3074
                     /gene "KIAA1233"
                     /note "Start codon is not identified."
                     /codon_start= 3
                     /product= "KIAA1233 protein"
                     /protein_id= "BAA86547.1"
                     /db_xref= "GI:6730729"
                     /translation= "AVCLEIQTQCTVNLSTLMVnsFAMSQACNIEECPPRWHVGSW
                     GPCSATCGVGTRDVYCLHPGETPAPLDECPDPKPHALQACNQFDCPPCWHIEEWQC
                     CSPTCGGGIQNRRVICRQLLIDGSFLNESPEICQGPKASSHKSCARTDCPPHLAVGDW
                     SPCSVSCGVGTQRRFQVCQRAAPGPRTPCLFMMCPPLLPGLPLVPSCQMPFCNKTFSS
                     MEIeCSpCpGyLvVQEVCvpTELERKLNETICEATLnPNIsVTIkCPVEsTQKSLI
                     QWEKPGPGIQNSKPRLGITKFSQSLKIHCIAAPDICVYFCIACSAQETVVLKLIGTPNRL
                     IARPALREPMREYPGMDHSFANSLGVTWHKMRQMWNNKNDLYLDDDHISNQPFLRALL
                     GHCSNSAGSTNSWELKNKQFEAAVKQCAYSMDTAQFDFLIRNMSQLMETGFVSDDLAS
                     QLIYQVVAEIAKAQPTHMQWPGIQPPTPPAAQLPCETQSVSQSSHAKNSGKLIFKPEG
                     PVLMRQSQPPSISFNKTINSPIGNTVYITKRTEVINILCDLITPSEATYTWTKDGTLL
                     QPSVKTILDGTGKIQIQNPTRKEQGIYECSVANHLGSDVESSSVLYAEAPVILSVERN
                     ITEFPSNHCQVVVQGITVEAALQANVTIPSFVKQVEQFSWKRQGSSLSCNVSLLHN
                     SLLLQNVSLPNPGTYVCIAINALCKAVAISVEHELPFPWPFSFIVFIQCHKKYIIQAI
                     NTKTNSNLPTGEFPPQEPFWEPGNWSHCSATCHEGARIQRFLCVMAHGQFVGEAICL
                     HLQKPLAGFEPQNIRDCPAPWITSVWSQCSVSCGCGYHSRQVTCKRTKANGTVQVVSP
                     KACAPKDKPLAGKKPCFGHPCVQWEPGNKCPGRCMGKAVRMQQRHTACQHNSSDSNCDD
                     PKMPTLREKNCTSGACDYCWHTGPWFPCTAACGPGFQSRPVVCIRTLPSCKPVAPRPCYQ
                     FPFTLPWRHCSLSCGVGFSTPRCMFVPRSNLDSCTFTF5SSLCLE"
BASE COUNT      1444 a    1062 c    1235 g    1366 t
ORIGIN
```

```
  Query Match             62.5%;  Score 3072.4;  DB 9;  Length 5107;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 3073;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy    2063 tagcagtgtgcttacatatccagagccagcagacagtcaatgacagcttgtgtgatatgg 2062
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 TAGCAGTGTGCTTACATATCCAGAGCCAGCAGACAGTCAATGACAGCTTGTGTGATATGG 60

Qy    2062 tccaccgtccctccagccatgagccaggccctgtaacacagagcccctgtccccccaggtggc 2122
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 TCCACCGTCCTCCAGCCATGAGCCAGGCCCTGTAACACAGAGCCCTGTCCCCCCAGGTGGC 120

Qy    2123 atgtgggctcttgggggccctgctcagctacctgtggagttggaattcagacccgagatg 2182
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 ATGTGGGCTCTTGGGGGCCCTGCTCAGCTACCTGTGGAGTTGGAATTCAGACCCGAGATG 180

Qy    2183 tgtactgcctgcagcccagtgagacccctgcccctcctgaggagtgccgagatgaaaagc 2242
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 TGTACTGCCTGCAGCCCAGTGAGACCCCTGCCCCTCCTGAGGAGTGCCGAGATGAAAAGC 240

Qy    2243 cccatgcttttacaagcatgccaatcagtttgactgcctcctggctggcacattgaagaat 2302
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 CCCATGCTTTACAAGCATGCCAATCAGTTTGACTGCCTCCTGGCTGGCACATTGAAGAAT 300

Qy    2303 ggcagcagtgttccaggagttggtggcgggggaactcagaacagaagagtcacctgtcggc 2362
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 GGCAGCAGTGTTCCAGGAGTTGGTGGCGGGGGAACTCAGAACAGAAGAGTCACCTGTCGGC 360

Qy    2363 agctgctaacggagtggcagctttttgaatctctcagatgaattgtgccaaggacccaagg 2422
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 AGCTGCTAACGGAGTGGCAGCTTTTTGAATCTCTCAGATGAATTGTGCCAAGGACCCAAGG 420

Qy    2423 catcgtctcacaaagtcctgtgccaggacagactgtcctccacatttagctgtgaggaact 2482
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 CATCGTCTCACAAGTCCTGTGCCAGGACAGACTGTCCTCCACATTTAGCTGTGAGGAACT 480

Qy    2483 ggtcgaaagtgttctgtcagttgtggtgttggaatccagagaagaaagcaggtgtgtcaaa 2542
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 GGTCGAAGTGTTCTGTCAGTTGTGGTGTTGGAATCCAGAGAAGAAAGCAGGTGTGTCAAA 540

Qy    2543 ggctggcagccaaaggtcggcgcatgcccctcagtgagatgatgtggcagggatctaccag 2602
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 GGCTGGCAGCCAAAGGTCGGCGCATGCCCCTCAGTGAGATGATGTGGCAGGGATCTACCAG 600

Qy    2603 agtttcctcttgtaaagatcttgccagatgccagtgagtaaaatccaatccagaagatga 2662
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 GGCTCCCTCTTGTAAGATCTTGCCAGATGCCTGAGTGCAGTAAAATCCAAATCAGAGATGA 660

Qy    2663 agacaaaacttgtgtgagcaagtgtccgcctgatgtccgcagtgtgtctgcatttcccaga 2722
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 AGACAAAACTTGGTGAGCAGGGTCCGCAGATGCTCAGTGTGTCAGAGAGTCTACATTCAGA 720

Qy    2723 cgagtgaagaagaagaagtgtattaagctgccccagtgcagcattccagaaatgtctgcat 2782
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 CGAGTGAAGAAGAAGAAGTGTATTAAGCTGCCCCAGTGCAGCATTCCAGAAATGTCTGCAT 780

Qy    2783 ccatccgtgattattaagtgccccagtgcgacgattccagaaatgtctgcatccaagtgaaga 2842
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 CCATCCGTGATTATTAAGTGCCCCAGTGCGACGATTCCAGAAATGTCTGCATCCAGTGAAGA 840

Qy    2843 agtagtgcgcagttgcctgcagaagctccagaaacgagcttagcgatccacgagtccagcct 2902
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 AGTAGTGCGCAGTTGCCTGCAGAAGCTCCAAACGAGCTTAGCGATCCACGAGTCCAGCCT 900

Qy    2903 aaatccacggtcttgctgcccccgacatcggcgtgtaccggtgccattgccagggctctgcac 2962
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 AAATCCACGGTCTTGCTGCCCCCGACATCGGCGTGTACCGGTGCCATTGCCAGGGCTCTGCAC 960

Qy    2963 agtttaaacaattgtgctcaagctcattggtactgacaaccggctcatcgccagccccagccc 3022
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 AGGAAACAGTTGTGCTCAAGCTCATTGGTACTGACAACCGGCTCATCGCCAGCCCCAGCCC 1020

Qy    3023 tcaggaggccctatgagggaatatcctggaatggaccacagcgaagccaatagtttgggag 3082
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 TCAGGGAGCCTATGAGGGAATATCCTGGAATGGACCACAGCGAAGCCAATAGTTTGGGAG 1080

Qy    3083 tcacatggcgacaaaatgaggcaaatgtggaaataacaaaaatgacctttatctggatgatc 3142
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 TCACATGGCACAAAATGAGGCAAATGTGGAATAACAAAAATGACCTTTATCTGGATGATC 1140

Qy    3143 accacattagtaaccagcctttcttgagagctctgttaggccactgcagcaattctgcag 3202
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 ACCACATTAGTAACCAGCCTTTCTTGAGAGCTCTGTTAGGCCACTGCAGCAATTCTGCAG 1200

Qy    3203 gaagcaccaactcctgtggagttgaagaataagcagtttgaagcagcagctgttaaacaaggag 3262
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 GAAGCACCAACTCCTGTGGAGTTGAAGAATAAGCAGTTTGAAGCAGCAGTTAAACAAGGAG 1260

Qy    3263 catatagcatggatacagcccagtttgatgagctgataagaaacatgagtcagctcatgg 3322
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1261 CATATAGCATGGATACAGCCCAGTTTGATGAGCTGATAAGAAACATGAGTCAGCTCATGG 1320

Qy    3323 aaaccggagaggtcagcgatgatcttgcgtcccagctgatatatcaggctggtggccgaat 3382
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1321 AAACCGGAGAGGTCAGCGATGATCTTGCGTCCCAGCTGATATATCAGCTGGTGGCCGAAT 1380

Qy    3383 tagccaaggcacaggcaacacacatgcagtggcgggcatccaggaagagacacctcctg 3442
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1381 TAGCCAAGGCACAGCCAACACACATGCAGTGGCGGGGCATCCAGGAAGAGACACCTCCTG 1440

Qy    3443 ctgctcagctcagagggcaaacagggagtgtgtccaaagctccgcatgcaaaaactcag 3502
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1441 CTGCTCAGCTCAGAGGGCAAACAGGGAGTGTGTCCCAAAGCTCCGCATGCAAAAACTCAG 1500

Qy    3503 gcaagctgacattcaagccgaaaggaccttgttctcatgaggcaaagccaacctcctcaa 3562
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1501 GCAAGCTGACATTCAAGCCGAAAGGACCTTGTTCTCATGAGGCAAAGCCAACCTCCTCAA 1560

Qy    3563 tttcatttaataaaacaataaattccaggattggaaatacagtataacattacaaaaagga 3622
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1561 TTTCATTTAATAAAACAATAAATTCCAGGATTGGAAATACAGTATACATTACAAAAAGGA 1620

Qy    3623 tagagtgtatcaatatatgtgtgtaaattattaaaaaaatgaaaaaaaaaatatatgtaa 3682
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1621 CAGAGGTCATCAATATACTGTGTGACCTTATTACCCCCAGTGAGGCCACATATACATGGA 1680
```



```
RESULT  11
AX235411
LOCUS       AX235411               5107 bp    DNA     linear   PAT 11-SEP-2001
DEFINITION  Sequence 58 from Patent WO0162928.
ACCESSION   AX235411
VERSION     AX235411.1  GI:15593526
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 5107)
  AUTHORS   Vernet,C.A., Fernandes,E., Shimkets,R.A., Macdonald,J. and
            Spaderna,S.K.
  TITLE     Polypeptides and nucleic Acids encoding same
  JOURNAL   Patent: WO 0162928 A 58 30-AUG-2001;
            Curagen Corporation (US)
FEATURES             Location/Qualifiers
     source          1..5107
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT     1445 a   1061 c   1235 g   1366 t
ORIGIN


  Query Match          81.5%;  Score 3076.8;  DB 4;  Length 5107;
  Best Local Similarity   99.7%;  Pred. No. 0;
  Matches 3072;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy    2003 tagaaatgtgctgcatcatcccaaccccaaaaaaatgaatgacaacctgtgtgtatgtg 2062
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 TAGCAGTGTGCTTACATATCCAGACCAGCAGCAGTCAATGCAGCTTGTGTGTATGTG 60

Qy    2063 tccacctgtctccaacctatgaccaggctgtatccaacaaaaccctatcccccaagtgac 2122
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 TCCACCGACTCCTCAGCATGACCAGCCTGTAACCACAGACCCCTGTCCCCCAGGTGAC 120

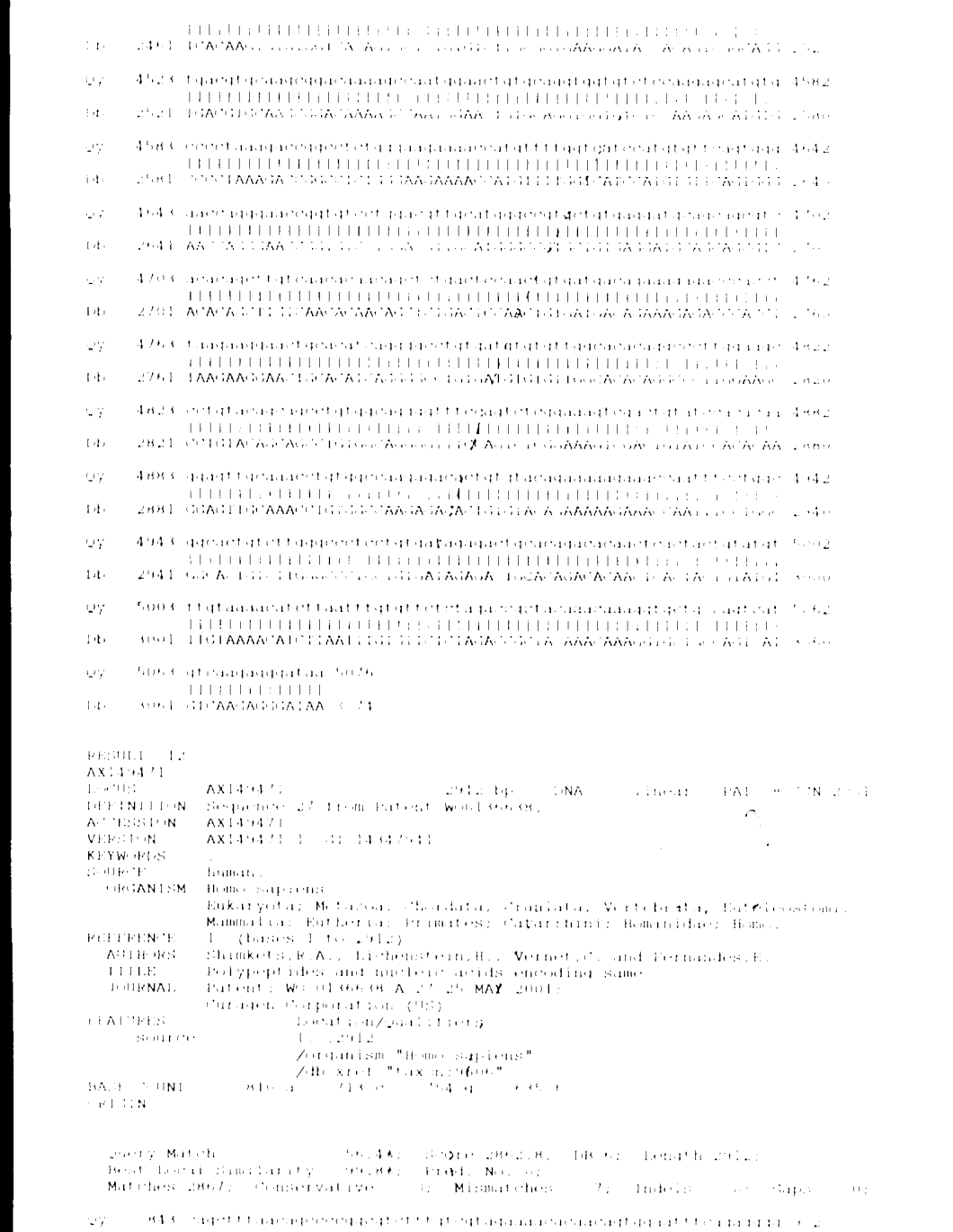
Qy    2123 atatgtgactctgaaaaccctgctgaggctgtgaagttgaaatccaaggaaactgatgaa 2182
           |||||||||||||||||||   ||||||||||||||||||||||||||||||||||||||
Db     121 ATGTGGCTCTTGGGGGCCTGCTACAGTACCTGTGGAGTTGGAATCCAGACCGACATG 180

Qy    2183 tgtactgcctgtgacccaaaaaaaaccctgcccctcctgaggaatgtccgaatgtaaaaac 2242
           |||||||||||||||||||| | ||||||||||||||||||||||||||||||||||||
Db     181 TGTACTGCTGCACCCAGGGAGACCCTGCCCCTCCAGGAGGAGTGCGAATGTAAAAAC 240

Qy    2243 ccaatgctttacaaacatgcaatgaatttgactgtcctcctgctgatgaatgtcaatttaatgaaat 2302
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 CCAATGCTTTACAAACATGCAATCAGTTTGACTGCCTCCTGGCTGGCATGATTGAAGAAT 300
```

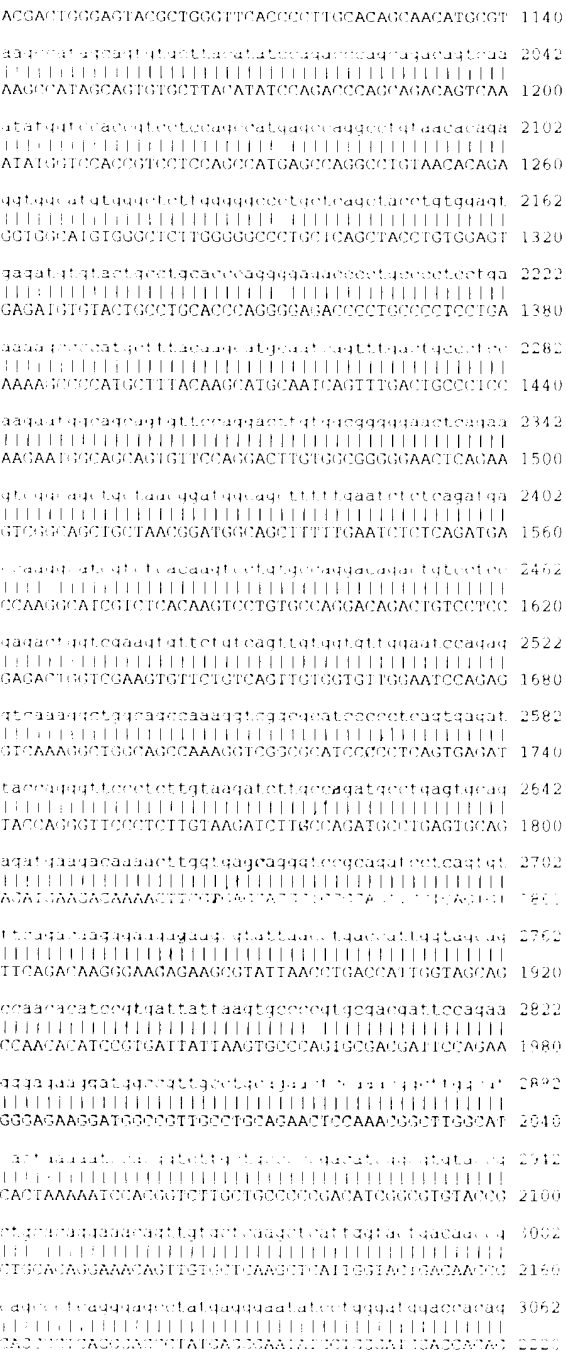


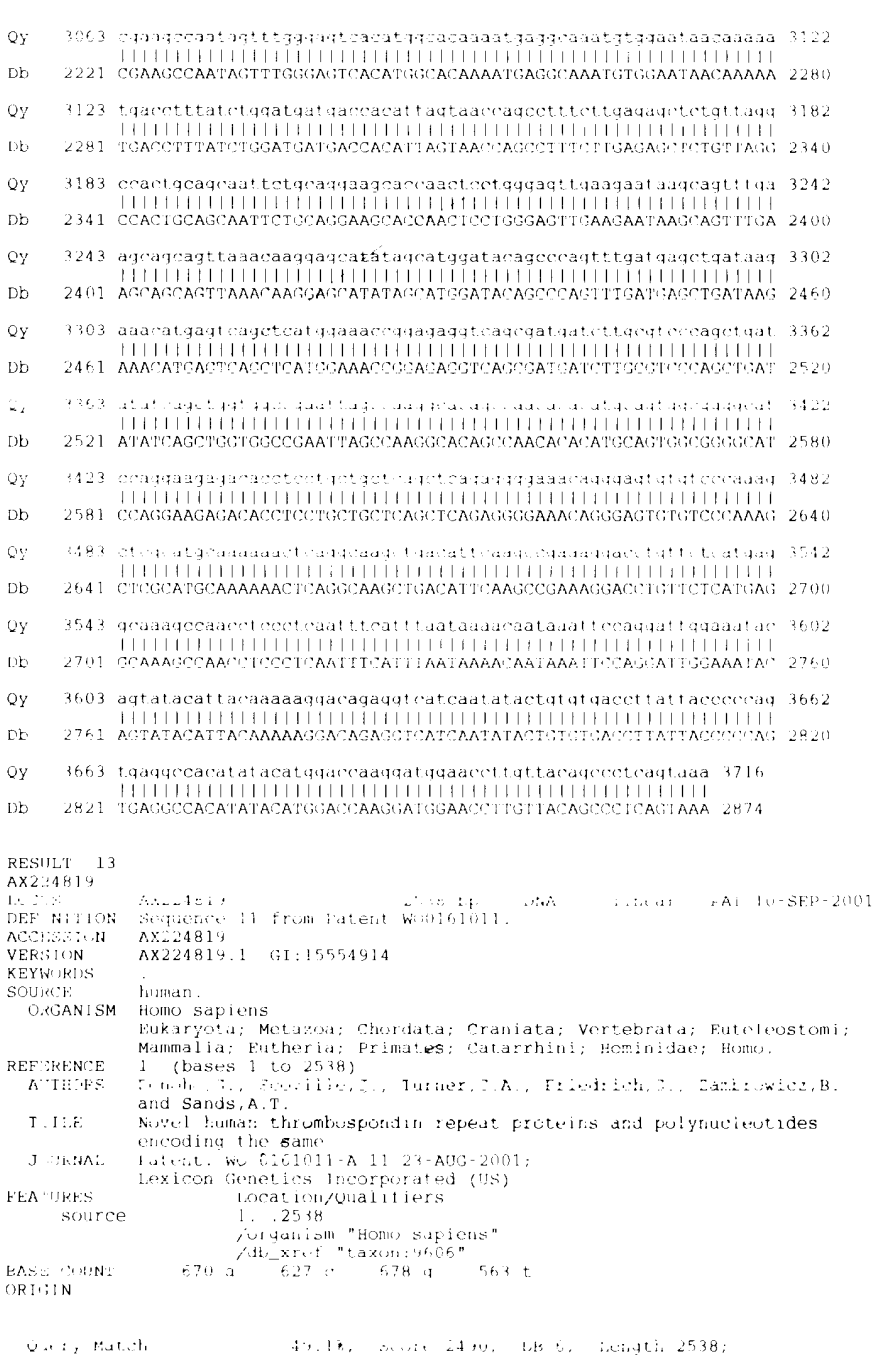




RESULT 12
AX149471
LOCUS AX149471 2912 bp DNA linear PAT 19-JUN-2001
DEFINITION Sequence 27 from Patent WO0136638.
ACCESSION AX149471
VERSION AX149471.1 GI:14467511
KEYWORDS .
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2912)
AUTHORS Shimkets,R.A., Lichenstein,H., Vernet,C. and Fernandes,E.
TITLE Polypeptides and nucleic acids encoding same
JOURNAL Patent: WO 0136638-A 27-25-MAY-2001;
Curagen Corporation (US)
FEATURES Location/Qualifiers
source 1..2912
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT [illegible]
ORIGIN

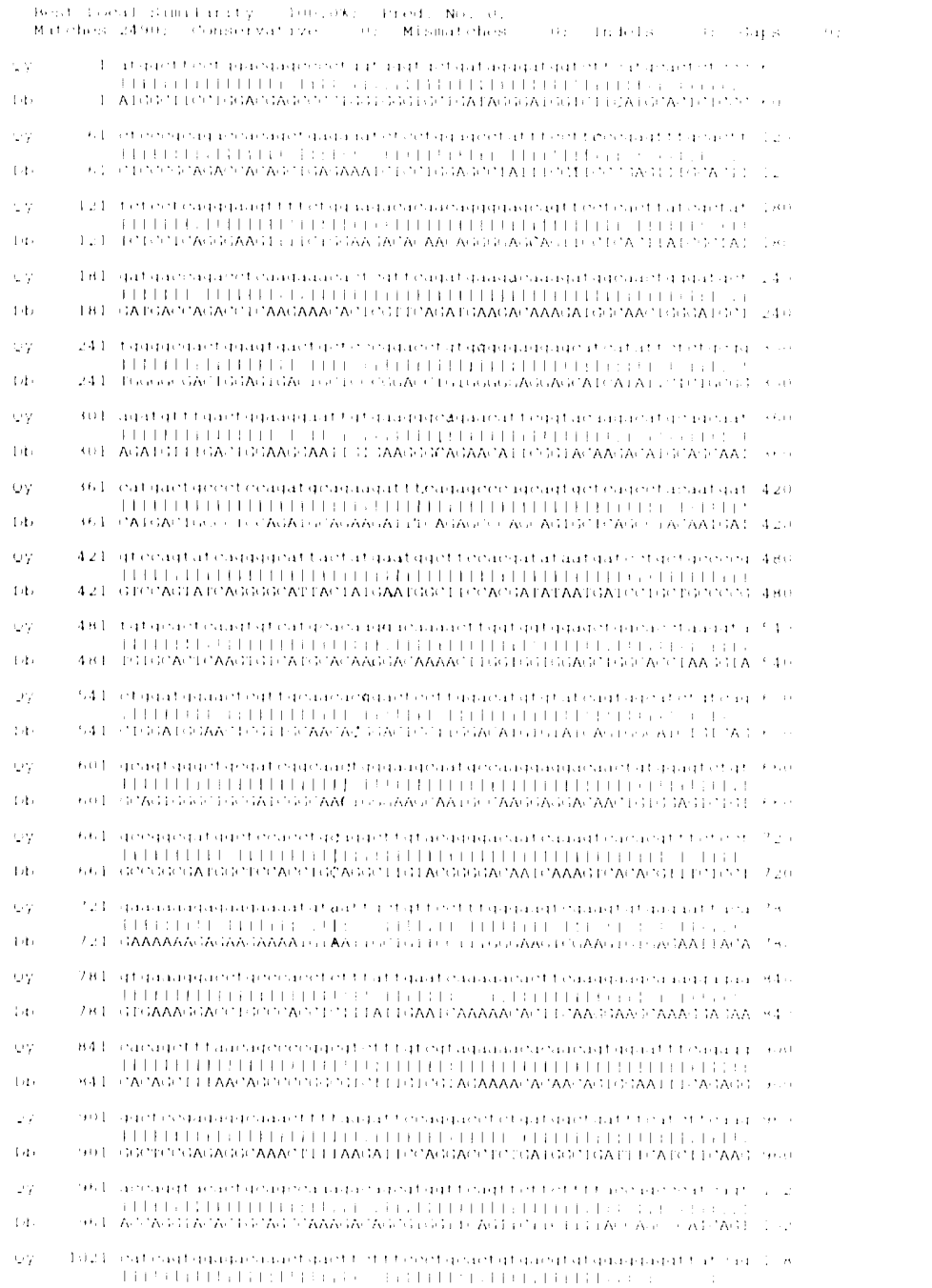
Query Match [illegible]; Score 2062.8; DB 9; Length 2912;
Best Local Similarity [illegible]; Pred. No. [illegible];
Matches 2067; Conservative [illegible]; Mismatches [illegible]; Indels [illegible]; Gaps [illegible];

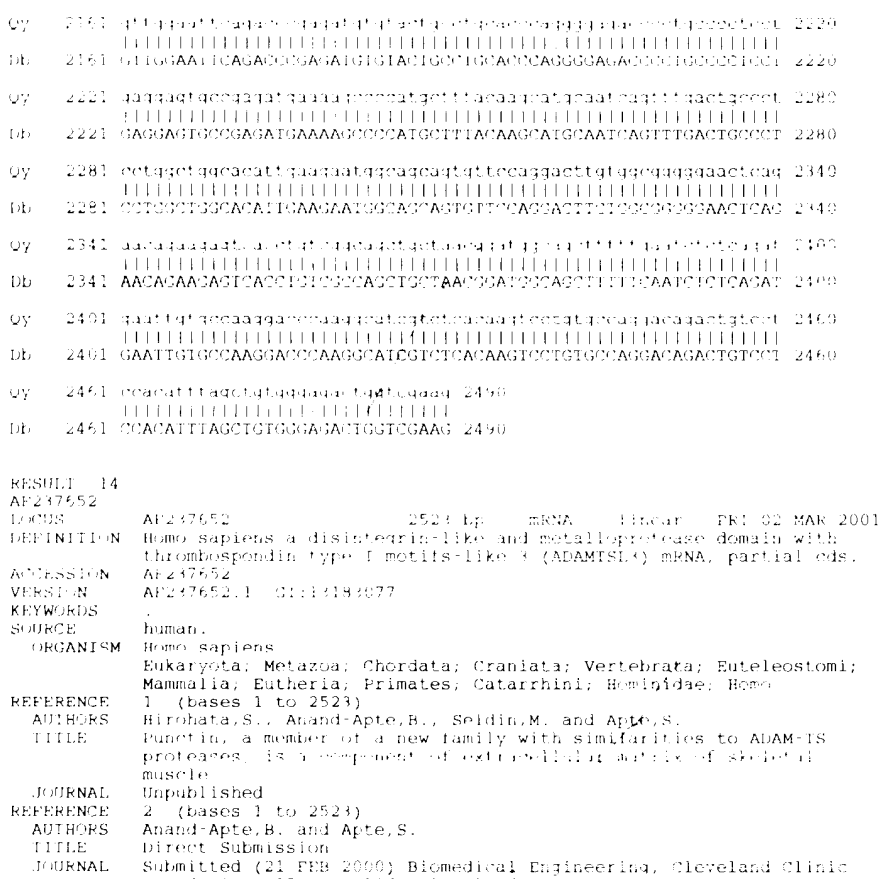




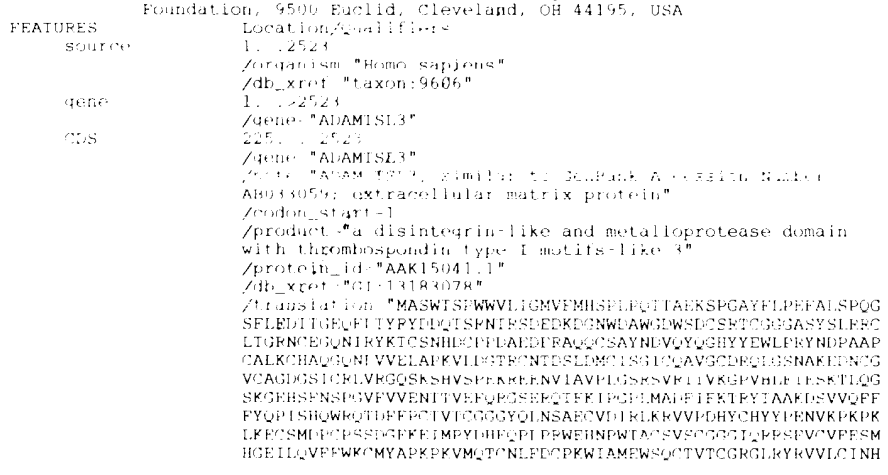
```
RESULT  13
AX224819
LOCUS       AX224819                2538 bp    DNA     linear   PAT 10-SEP-2001
DEFINITION  Sequence 11 from Patent WO0161011.
ACCESSION   AX224819
VERSION     AX224819.1  GI:15554914
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2538)
  AUTHORS   Tumb,T., Pacifici,S., Turner,C.A., Friedrich,G., Zamirowicz,B.
            and Sands,A.T.
  TITLE     Novel human thrombospondin repeat proteins and polynucleotides
            encoding the same
  JOURNAL   Patent: WO 0161011-A 11 23-AUG-2001;
            Lexicon Genetics Incorporated (US)
FEATURES             Location/Qualifiers
     source          1..2538
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      670 a    627 c    678 g    563 t
ORIGIN


  Query Match           45.1%;  Score 2490;  DB 6;  Length 2538;
```

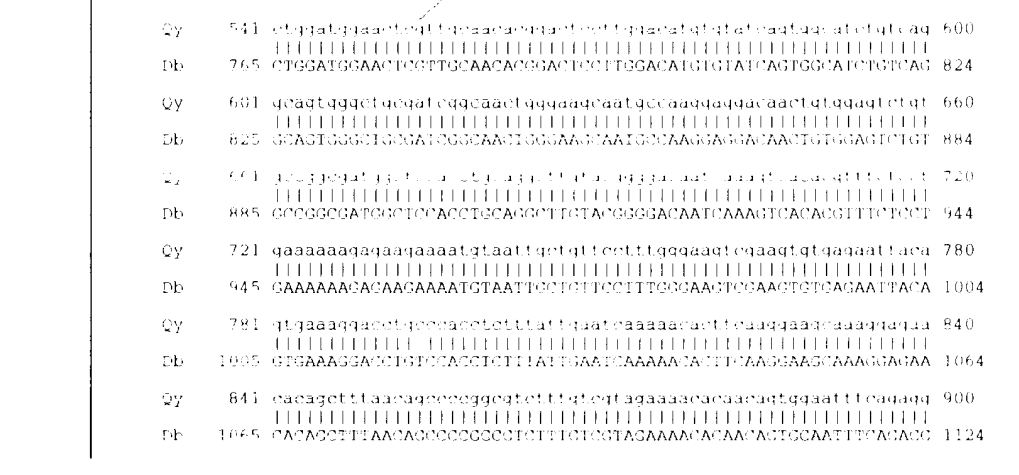




RESULT 14
AF237652
LOCUS AF237652 2523 bp mRNA linear PRI 02 MAR 2001
DEFINITION Homo sapiens a disintegrin-like and metalloprotease domain with thrombospondin type I motifs-like 3 (ADAMTSL3) mRNA, partial cds.
ACCESSION AF237652
VERSION AF237652.1 GI:13183078
KEYWORDS .
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo
REFERENCE 1 (bases 1 to 2523)
AUTHORS Hirohata,S., Anand-Apte,B., Seldin,M. and Apte,S.
TITLE Punctin, a member of a new family with similarities to ADAM-TS proteases, is a component of extracellular matrix of skeletal muscle
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2523)
AUTHORS Anand-Apte,B. and Apte,S.
TITLE Direct Submission
JOURNAL Submitted (21 FEB 2000) Biomedical Engineering, Cleveland Clinic Foundation, 9500 Euclid, Cleveland, OH 44195, USA
FEATURES Location/Qualifiers
source 1..2523
/organism="Homo sapiens"
/db_xref="taxon:9606"
gene 1..2523
/gene="ADAMTSL3"
CDS 225..2523
/gene="ADAMTSL3"
/note="ADAM-TS-like; similar to GenBank Accession Number AB033059; extracellular matrix protein"
/codon_start=1
/product="a disintegrin-like and metalloprotease domain with thrombospondin type I motifs-like 3"
/protein_id="AAK15041.1"
/db_xref="GI:13183078"
/translation="MASWTSPWWVLIGMVFMHSPLFQTIAEKSPGAYFLPEFALSPQG
SFLEDIIGEQFLIYPYPDQISPNIESDEDKDGNWDAWGDWSDCSRTCGGGASYSLEFC
LTGRNCEGQNIRYKTCSNHECPPLAIDFRAQQCSAYNEVQYQGHYYEWLPRYNDPAAP
CALKCHAQGQNIVVELAPKVLEVTPCNTDSLDMQISGIQGAVSCDEQLDSNAKEFNCG
VCAGDGSICRLVRGQSKSHVSPEKREENVIAVPLGSKSVRTLVKGPVHLETESKTLQG
SKGEHSFNSPGVFVVENITVEFQRQSEPQTFKIPCFLMAIFIEKTPYIAAEDSVVQPF
PYQPISHQWRQTDPFFPCTVTCGGGYQLNSAECVDIRLKRVVPDHYCHVYVENVKPKPK
LKECSMDPCPSSPGPFEFIMPYDHFQPIPPWFHNPWTAVSVSGGGIQPPSFVCVFESM
HGEILQVPFWKCMYAPKPKVMQTCNLFDCPKWIAMEWSQCTVTCGRGLKYRVVLCINH
RGEHVGGCNPQLKPIKKFPCVIPIPCYKPKFKSPVFAKIPWIKQAQELEETRIATEEP
IPIPEPWSACSITCGRGVQVREVECPRVLIFIQIETELPEEFCPGPELPTERPCLLEA
CDESPASRELDIPLPEDSFTTYDWEYAGFTPCTATCLGGHQEAIAVCLRIQTQQTVND
SLCDMVERPEAMSQACNAEDCPKWHVGSWGPCSATCGVGIQTRDVYCLHPGETPAPP
EECRDEKPHALQACNQFDCPPGWHIE"
BASE COUNT 635 a 652 c 692 g 544 t
ORIGIN



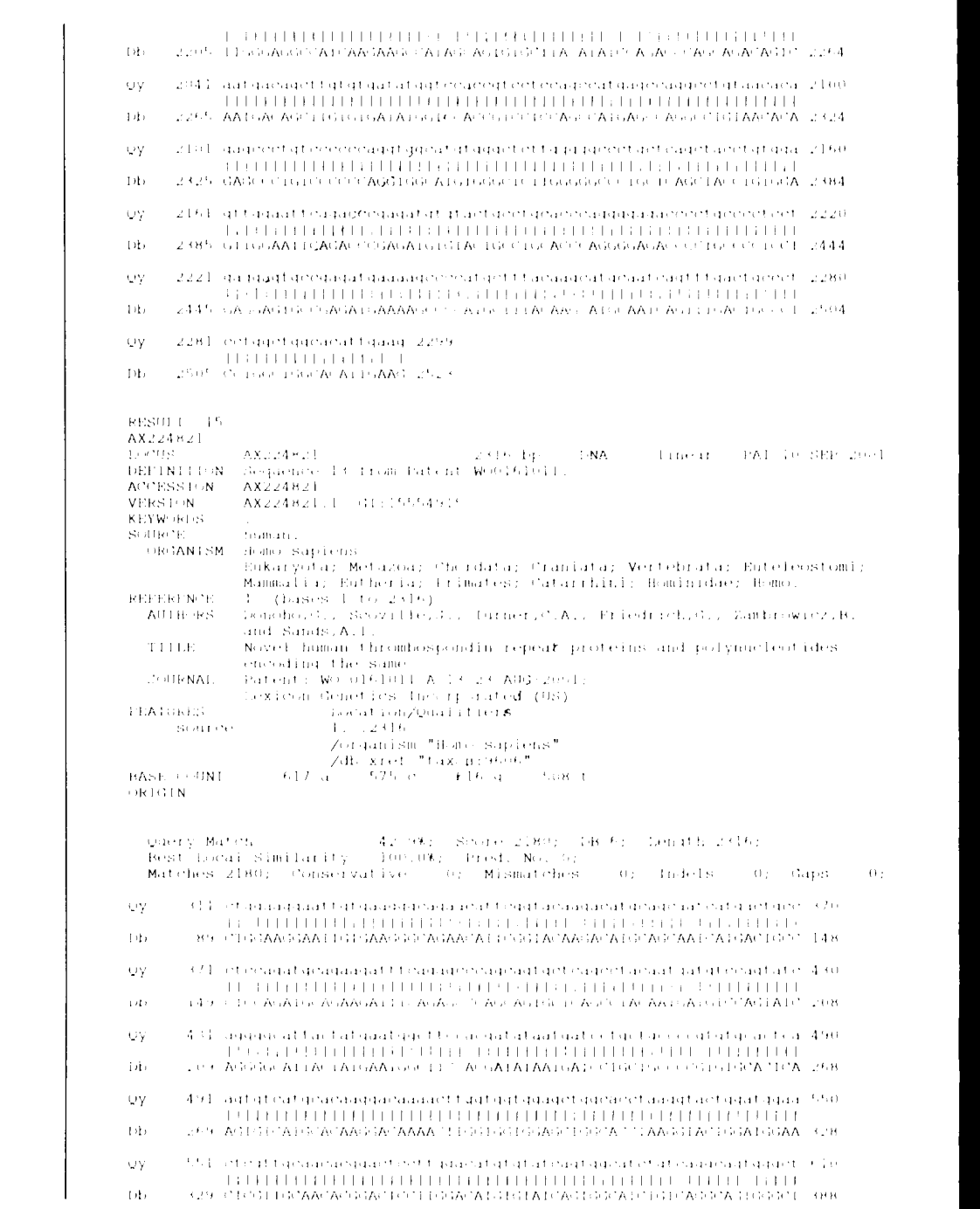
Query Match 45.2%; Score 2295.8; DB 9; Length 2523;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2297; Conservative 0; Mismatches 2; Indels 0; Gaps 0;





RESULT 15
AX224821
LOCUS AX224821 2516 bp DNA linear PAT 10-SEP-2001
DEFINITION Sequence 14 from Patent WO0161011.
ACCESSION AX224821
VERSION AX224821.1 GI:15554913
KEYWORDS .
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2516)
AUTHORS Donohoe,D., Scozzi,S., Turner,C.A., Friedrich,G., Zambrowicz,B.
and Sands,A.T.
TITLE Novel human thrombospondin repeat proteins and polynucleotides
encoding the same
JOURNAL Patent: WO 0161011 A 23-AUG-2001;
Lexicon Genetics Incorporated (US)
FEATURES /organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 617 a 525 c 416 g 958 t
ORIGIN

Query Match 42.9%; Score 2080; DB 6; Length 2516;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;







Search completed: August 18, 2002, 05:45:23
Job time: 10213 sec

was cleaved with BamHI and XhoI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda ZAP I."
BASE COUNT 161 a 163 c 191 g 145 t
ORIGIN

Query Match 5.2%; Score 262.4; DB 9; Length 660;
Best Local Similarity 70.6%; Pred. No. 1.4e-58;
Matches 350; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

Qy 186 ccaaqaccctcaaqaaacaactcqttcaqatqaaaqacaaaqatqqccaactqqqatqcttqqaq 245
Db 164 CCTGAGTTCCAGGACTGCAGCTCAGAGGAAGACCGTGAAGGCCTCTGGGATGCCTGGGG 223

Qy 246 cqactqqaaqtqactqctcccqaacttqtqqqqaqqaqcatcatattctctqccaqaatq 305
Db 224 CCCTTGGAGCGAGTGTTCCAGAACTGTGGTGGGGGTGCCTCCTATTCCCTGAGACGCTG 283

Qy 306 tttqactqqaaaqaattqtqaaqqqcaqaacattcqqtacaaqacatqcaqcaatcatqa 365
Db 284 CCTGAGCAGCAAGAGCTGTGAAGGAGAAATATTCGATATAGAACATGCAGTAATGTGGA 343

Qy 366 ctqcccccaaatqcaaaaattcaqaqcccaqcaqtqctcaqcctacaattatqtcca 420
Db 344 CTGCCCACCAGAAGCAGGTGATTCCGAGCTCAACAGTGTTCTGCTCATAACGACGTCAA 403

Qy 426 qtatcaqqqcattactatqaaatqactcaccaqatataatqatcctqctqcccctqtqc 485
Db 404 GTACCATGGACAGCTTTATGAATGGCTTCCTGTATCTAATGACCCCGACAATCCATCTCC 463

Qy 486 actcaqtqtcatqcacaqqacaaaactqtqtqtqqqctqtcaactaaqatactqta 540
Db 464 ACTCAAGTGCCAAGCAAAAGGAACCAGCCTGGTTGTGGAACTAGCACCGAAGGTCTTAGA 523

Qy 546 tqqaactcqttqcaacacqqcttcttqqacatqtqtatcaqtqqcatctqtcaqqcaqt 605
Db 524 TGGTACCGGATGCTATACACAATCACTGGATATGTGCATCAGCGGCCTATGCCAAATTGT 583

Qy 606 qqqctqcqatqqqcaactctttaaqcqatqtccaaaqaqqacaactqtqtqatctqtqccqa 665
Db 584 TGGCTGTGATACCACACTGGGACCACTGTAAAGAAGATAACTGTGGAATGTGCAACGG 643

Qy 666 cqatqqctccaacctqc 681
Db 644 AGATGGCTCAACCTGC 659

RESULT 14
BF543337/c
LOCUS BF543337 408 bp mRNA linear EST 11-DEC-2000
DEFINITION UI-R-Y0-abq-a-03-0-UI.r1 UI-R-Y0 Rattus norvegicus cDNA clone UI-R-Y0-abq-a-03-0-UI 5', mRNA sequence.
ACCESSION BF543337
VERSION BF543337.1 GI:11644444
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 408)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 9250
Fax: 319 335 9555
Email: msoares@blue.weeg.uiowa.edu
cDNA Library Preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com) This clone is also available through the I.M.A.G.E. Consortium at LLNL (info@image.llnl.gov). IMAGE ID= 1790944
Seq primer: M13 Forward.
FEATURES Location/Qualifiers
source 1..408
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-Y0-abq-a-03-0-UI"
/clone_lib="UI-R-Y0"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT7T3-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-Y0 library is a subtracted library derived from an individually-tagged normalized whole-eye (minus the lens) library. The driver for the subtraction consisted of a pool of all previous libraries (UI-R-A0, UI-R-A1, UI-R-E0, UI-R-E1, UI-R-F0, and UI-R-C1). The tag is a string of 3'-5' nucleotides present between the Not I site and the oligo dT track which allows identification of the library of origin of a clone within the mixture. The subtracted library (UI-R-Y0) was constructed as follows: PCR amplified cDNA inserts from previous library clones from which 3' ESTs had been derived were used as a driver in a hybridization with the normalized whole eye library in the 2 rounds of single stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the UI-R-Y0 library. This procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996)."
BASE COUNT 92 a 113 c 107 g 99 t
ORIGIN

Query Match 7.1%; Score 359.6; DB 10; Length 408;
Best Local Similarity 85.5%; Pred. No. 6.5e-58;
Matches 289; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

Qy 1 atqqcttcctqaaqaaqcccctqqtqtttqctqataaqqatqqtqttcatqcaqcctctccc 60
Db 348 ATGCCTTCTGGAAACTATCTCCCCCTGTACACGAATGATCTTTACAGTCTCTGGT 289

Qy 61 ctcccqcaaqcaqcaqtttaaaat ttcctcqaqqcctatttcctttcccaqtttqcactt 120
Db 288 CTCTGCAGCCACACCAGAGAACTCCTGGAGCCTACTTCCTTCCTGAGTTCGCACTC 229

Qy 121 tctcctcaqqqaqtttctctqtaaaaqcaqcaqqqqqcaqttcctcactatqtqctat 180
Db 228 TCTCCTCAGGAAGCTTCAGAAGATACACAGGGAACAGTTCTCACATCCGCTAT 169

Qy 181 qatqaccaqaccctcaaqaaaaatctttcaaatqaaqacaaaqatqqccaactqqqatqct 240
Db 168 GATGATCAGACTGAAGAAATCCACTCCCATGAACAAAGATGGCAACTGGGATGCC 109

Qy 241 tqqqqaactqqaatqactqctcccqaacttqtqqqqqqaqcatcatattctctqccaq 300
Db 108 TGGGGAACTGGAGTCACTCTCCCGAACTGTGGGGAGCAGCCTCCTATTCTTTGCGA 49

Qy 301 aaatqtttqactqaaaaaattcttcaaaaaaaqaacat 338
Db 48 ACATGTCTGACTGGAAGAAGTCGAAGGACAGAACAT 11

RESULT 15
BI523618
LOCUS BI523618 581 bp mRNA linear EST 29-AUG-2001
DEFINITION yx69f12.r1 NIH_MGC 122 Homo sapiens cDNA clone IMAGE:2201199 5',

was cloned with BamHI and XhoI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda ZAP II."
BASE COUNT 161 a 164 c 191 g 145 t
ORIGIN

Query Match 5.2%; Score 262.4; DB 9; Length 660;
Best Local Similarity 70.6%; Pred. No. 1.4e-58;
Matches 450; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

Qy 186 ccaaacctcaaaaaaacatcagttcaaatgaaaacaaagatggcaactgagatgcttgggg 245
Db 164 CCTGAGTTCCAGGACTGCACCTCAGAGGAAGACCGTGAAGGCCTCTGGGATGCCTGGGG 223

Qy 246 cgactgaattaactgctcccgaactgtgggggaggtgctatattctctgcgggatg 305
Db 224 CCCTTGGAGCGAGTGTTCCAGAACTGTGGTGGGGGTGCCTCCTATTCCTGAGACGCTG 283

Qy 306 tttgactgggaaggaattgtgaaggccaaacattcggtgtcaagacatgcaagcaatcatga 365
Db 284 CCTGACCAGTAAGACCTGTGAAGGGAGAAATATTCGATATCAACATCCAGTAATGTGGA 343

Qy 366 ctgcccctcaagatgcaaaaaatttcaaaagcccaagcagtgctcaagcctacaaatgatgtcca 425
Db 344 CTGCCCACTCAGAAGCAGGTGATTCCCGAGCTCAACGTGTTCTGCTCTATAAGCAGATTCAA 403

Qy 426 gtatcagggcaattactatgaaatgacttcaagatatgaatgatcctgctgcccctgtgc 485
Db 404 GTACTATGGACACCTTATGAAACCCTCCTGTATTAATGACCGACAATCTATTCTC 459

Qy 486 actcaagtgtcatgcacaatgacaaaacttgttgagtgagctgcacctaagatgctgga 545
Db 460 ACTCAAGCCCAAGCAAAAGGAACAGCAGTGGGTGGAACTACCACTGAAGGTCTTAGA 519

Qy 546 tgaaactcgttgcaaacagcactcttgaacatgtgtatcaatgcgcatctgtcaggcagt 605
Db 520 TGCTACTGGATGCTATACAGAATACTGGATATGTGCATCAGCGGCTATGCTAAATGTGT 579

Qy 606 agcctcagcatcagcgcaactcgaaagcgcatgccaaggagcgcaactgtgagctgtgtgccgg 665
Db 580 DGGCTGTGATACCAACTGCAGCTGCTGTCAAAGAAGATAATGTGCAATGTGCAACGCT 639

Qy 666 cgatgggtccaacctgc 681
Db 640 AGATGGCTCAACTGCTC 656

RESULT 14
BF543337/c
LOCUS BF543337 408 bp mRNA linear EST 11-DEC-2001
DEFINITION UI-R-Y0-abq-a-03-0-UI.r1 UI-R-Y0 Rattus norvegicus cDNA clone UI-R-Y0-abq-a-03-0-UI 5', mRNA sequence.
ACCESSION BF543337
VERSION BF543337.1 GI:11934444
KEYWORDS EST
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 408)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319-335-8250
Fax: 319-335-9570
Email: msoares@blue.weeg.uiowa.edu
cDNA Library Preparation: M.B. Soares Lab clone distribution: clones will be available through Research Genetics (www.resgen.com) this clone is also available through the I.M.A.G.E. Consortium at LLNL (info@image.llnl.gov). IMAGE ID= 1790944
Seq primer: M13 Forward.
FEATURES Location/Qualifiers
source 1..408
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-Y0-abq-a-03-0-UI"
/clone_lib="UI-R-Y0"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT7T3-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-Y0 library is a subtracted library derived from an individually tagged normalized whole-eye (minus the lens) library. The driver for the subtraction consisted of a pool of all previous libraries (UI-R-A0, UI-R-A1, UI-R-E0, UI-R-E1, UI-R-C0, and UI-R-C1). The tag is a string of 336 nucleotides present between the Not I site and the oligo dT track which allows identification of the library of origin of a clone within the mixture. The subtracted library (UI-R-Y0) was constructed as follows: PCR amplified cDNA inserts from previous library clones from which 3' ESTs had been derived were used as a driver in a hybridization with the normalized whole eye library in the 2 cot 4 phase stranded circles. The remaining unhybridized circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the UI-R-Y0 library. The procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996)."
BASE COUNT 91 a 114 c 104 g 99 t
ORIGIN

Query Match 5.2%; Score 262.4; DB 9; Length 408;
Best Local Similarity 70.6%; Pred. No. 1.4e-58;
Matches 289; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

Qy 1 gtaaacctcaaaaaaacatcagttcaaatgaaaacaaagatggcaactgagatgcttgggg 60
Db 448 ACTGAGTTCCAGGACTGCACCTCAGAGGAAGACCGTGAAGGCCTCTGGGATGCCTGGGG 389

Qy 61 cgactgaattaactgctcccgaactgtgggggaggtgctatattctctgcgggatttgcaactt 120
Db 388 CCTTGGAGCGAGTGTTCCAGAACTGTGGTGGGGGTGCCTCCTATTCCTGAGTTCGCACTTC 329

Qy 121 tgtgctcctgaaaagttttctgaaaagcaagcaaccgaaaagcaattctgctcagttatgcagctat 180
Db 328 TCTCCTCAGGGAAGCTTTCTAGAAGATACCACAGGGGAACAGTTCCTCACATACCGCTAT 269

Qy 181 gatgagccagaacctcaaaaaaacatcagttcaaatgaaaacaaagatggcaactgagatgcttgggt 240
Db 268 GATGATCAGACCTCAAGAAATGCTCACTCGGATGAAGACAAAGATGGCAACTGGGATGCCTGGGT 209

Qy 241 tgactgaactgaagtgactgtgcccgaacctatgggggaggtgctatattctctgcgggg 300
Db 208 TGGGGAGACTGGAGTGACTGTCCCGACCTGTGGGGAGGAGCCTCCTATTCCTTTGTGGA 149

Qy 301 cgtatttgactgaaaaaaatttatgaaaaaaaacat 338
Db 148 ACATACCTTGACTTGAAGCAACTCTGAGGACAGAACAT 111

RESULT 15
BI523618
LOCUS BI523618 581 bp mRNA linear EST 29-AUG-2001
DEFINITION IMAGE:5201199 5', mRNA sequence.

Taqami,M., Taqawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
JOURNAL Unpublished (2001)
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9210
Email: genome-te@gsc.riken.go.jp
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
Kawai,K., Fujiwara,S., Inoue,K., Ishiwa,S., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,I., Sawa,K., Tanaka,T., Matsushita,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.
Computer based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Kondo,S., Shimadaira,A., Saito,R., Kiyosawa,H., Yamanaka,I., Arakawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences. Mamm. Genome 12, 654-657 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.
6 mouse tissues.

FEATURES Location/Qualifiers
source 1..622
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="A530027F04"
/clone_lib="RIKEN full-length enriched, adult male spinal cord"
/sex="male"
/tissue_type="spinal cord"
/dev_stage="adult"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI. cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in RIKEN Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in RIKEN contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAACAGCTCTTTTTTTTTTTTTTTVN 3']. cDNA was prepared by using trehalose thermoactivated reverse transcriptase and subsequently enriched for full length by cap trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 450.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAA]. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

BASE COUNT 151 a 157 c 172 g 142 t
ORIGIN

Query Match 5.8%; Score 293.6; DB 9; Length 622;
Best Local Similarity 89.1%; Pred. No. 6.5e-62;
Matches 401; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

```
Qy      186 ccagacctcaagaaaaractnattsaqatqaaqucaaaqatqqcaactqqqatqcttqqqq 245
            || ||  || || |   | || ||||| ||||||   || ||| |||||||| ||||||
Db       25 CCTGAGTTCCAGGACTGCACGCTCAGAGGAAGACCGTGAAGGCCTCTGGGATGCCTGGGG 84

Qy      246 cqactqqaqtqactqctccqqqacctqtqqqqaaqaqcatcatattctctqccaqaatq 305
            |   ||||| || || ||| | |||||||| || || || || ||||| ||| |  | ||
Db       85 CCCTTGGAGCGAGTGTTGCACAACCTGTGGTGAGGGTGCCTCTATTCCTGACACGCTG 144

Qy      306 tttqactqqaaqqaattqtqaaqqccaqaacattcqqtacaaaqacatqcaqcaatcatqa 365
            |||  | | ||  |||||||||   || ||||| || | |||||||| |||   ||
Db      145 CTGAGCAGCAAGAGCTGTGAAGGAGAAATATTCGATATACAACATGCAGTAATGTGCA 204

Qy      366 ctqcctccaqatqcaqaaqatttcaqaatccaatcaqtqctcaqcctacaatqatqtccq 425
            |||||| ||||| |||  |||||| |||| || ||||  || |||| | ||| || |||| |
Db      205 CTGCCCACCAGAAGCAGGTGATTTCCGAGTTCAACAGTGTTCTGCTCATAACCGACGTCAA 264

Qy      426 qtatcaqqqqcattactatqaaatqqcttccacqatataatqatcctqctqccccqtqtqc 485
            ||| || || ||   |||||||||||||||   || |||||| || |     || || |
Db      265 GTACCATGGACAGCTTTATGAATGGCTTCCTGTATCTAATGACCCCGACAATCCATGCTC 324

Qy      486 actcaaqtqtcatqcacaaqqacaaaactttqqtqqtqqaactqqcacctaaqqtactqqa 545
            |||||||||| || ||| ||||   | | |||| |||||| || ||||| |||||  | ||
Db      325 ACTCAAGTGCCAAGCAAAAGGAACCAGCCTGGTTGTGGAACTAGCACCGAAGGTCTTAGA 384

Qy      546 tqqaactcqttqcaaacacqqactcctttqqacatqtatatcaqtqucatctqtcaqqcaqt 605
            ||| || || ||| | |  || ||  |||| ||||| ||||| || || || ||  ||
Db      385 TGGTACCGCGATGCTATACAGAATCACTGGATATGTGCATCAGCGGCCTATGCCAAATTGT 444

Qy      606 qqactqcqatatcqqcaactttqqaaqcaatqccaqqqaqqaqacaactatqqaqtctqtqccqq 665
            |||||  ||||   || |||||  |||| || ||| |||||  ||||||||||| ||   |||
Db      445 TGGCTGTGATCACCAGCTGGGGAGCACTGTCAAAGAGGATAACTGTGGAGTGTGCCACGG 504

Qy      666 cqatqqctccacctqcaqqcttqtacqqqqacaatcaaaqtcacacqtttctcctqaaaa 725
            |||||||| || ||| |||| || || || || ||  || | |||| | ||||  ||
Db      505 AGATGGCTCAACGTGCCGCCTGTCCGAGGGCAGTATAAATCTCAGCTCTCTGCAACTAA 564

Qy      726 aaqaaaaaqaaaatqtaattqctqttcctttqqaaaqtcqa 765
            |  || || |  ||  | ||  ||||||  || ||  ||
Db      565 ATGGATGACACCGTGGTCGCCATTCCTTACGGCAGCAGA 604
```

RESULT 8
AQ743958
LOCUS AQ743958 890 bp DNA linear GSS 16-JUL-1999
DEFINITION BS_5504_A1_D01_T7A RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate=1080 Col=1 Row=G, DNA sequence.
ACCESSION AQ743958
VERSION AQ743958.1 GI:5521480
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 890)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)

```
            or from Research Genetics (info@resgen.com). BAC end Web Server:
            http://www.hgsc.bcm.tmc.edu
            Plate: 1080  row: G  column: 1
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 890.
FEATURES             Location/Qualifiers
     source          1..890
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate:1080 Col-1 Row:G"
                     /clone_lib "RPCI-11 Human Male BAC Library"
                     /sex="male"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     Male blood DNA was isolated from one randomly chosen donor
                     and partially digested with a combination of EcoRI and
                     EcoRI Methylase. Size selected DNA was cloned into the
                     pBACe3.6 vector at EcoRI sites"
BASE COUNT       228 a    185 c    218 g    259 t
ORIGIN


  Query Match          5.6%;  Score 281.8;  DB 12;  Length 890;
  Best Local Similarity   97.6%;  Pred. No. 1.1e-64;
  Matches  286;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy    1696 ccaaacgttcattccagaaccctgttcaacctgcattacaccatgtgagccaagtatgcaa 1755
           || | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     125 CCTAGGTTCATTCCAGAACCCTGTTCAGCTGCATTACACCATGTGAGCCAAGTATGCAA 184

Qy    1756 gtccgcgcagatgcaagcgtcgtgtgtctcacattcacgcatgtaagtaagctgcgcc 1815
           ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     185 TCCCGTCAGCTGAAGCTCCGTCTCTCACATTCACGCAGACTGATTATTCGTCTCCCC 244

Qy    1816 gagagaagtgtgaagcccccaagctgccgcgaagcgcgctgctcgtaaaagcatgt 1875
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     245 GACTAAGACTTCTGAAGCGCCAACTTCTCCACGAACGCCCTCGTCTCTTGAAACTACTGT 304

Qy    1876 gatcgagagccccgcgctccccgaaagctagacatcccgtcgcgaaagcaatgagacgagt 1935
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     305 GATCAGCAGCCCCGGCTCCCGAGAGCTAGACATCCCTCTCCTGAGGACAGTCGAGACGAC 364

Qy    1936 tacgagctgcgagtacgctgagttcaccttgcacgagcaagcatgctgagtcgagtcgtgt 1995
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     365 TACGACTCGCCAGTACGCTCGCTTCACCCTTGCACAGCAACATGCGTCGCGACGCGT 417


RESULT  9
BE888902
LOCUS       BE888902                 850 bp    mRNA    linear   EST 25-OCT-2000
DEFINITION  601513828T1 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:3915294 5',
            mRNA sequence.
ACCESSION   BE888902
VERSION     BE888902.1  GI:10345670
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 850)
  AUTHORS   NIH MGC Project
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9738  row: a  column: 07
            High quality sequence stop: 742.
FEATURES             Location/Qualifiers
     source          1..850
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3915294"
                     /clone_lib="NIH MGC 71"
                     /tissue_type "leiomyosarcoma"
                     /lab_host "LIBH06 (phage-resistant)"
                     /note= " Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primers: Oligo dT.
                     Average insert size 2.1 kb.  "
BASE COUNT       [illegible]
ORIGIN


  Query Match          5.2%;  Score 278;  DB 10;  Length 850;
  Best Local Similarity   97.4%;  Pred. No. 1.1e-62;
  Matches  451;  Conservative    2;  Mismatches    2;  Indels    4;  Gaps    4;

Qy     186 ccaaacgttcattccagaaccctgttcaacctgcattacaccatgtgagccaagtatgcaa 245
           || | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     103 CCTAGGTTCATTCCAGAACCCTGTTCAGCTGCATTACACCATGTGAGCCAAGTATGCAA 162

Qy     246 gtccgcgcagatgcaagcgtcgtgtgtctcacattcacgcatgtaagtaagctgcgcc 305
           ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     163 TCCCGTCAGCTGAAGCTCCGTCTCTCACATTCACGCAGACTGATTATTCGTCTCCCC 222

Qy     306 gagagaagtgtgaagcccccaagctgccgcgaagcgcgctgctcgtaaaagcatgt 365
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     223 GACTAAGACTTCTGAAGCGCCAACTTCTCCACGAACGCCCTCGTCTCTTGAAACTACTGT 282

Qy     366 gatcgagagccccgcgctccccgaaagctagacatcccgtcgcgaaagcaatgagacgagt 425
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     283 GATCAGCAGCCCCGGCTCCCGAGAGCTAGACATCCCTCTCCTGAGGACAGTCGAGACGAC 342

Qy     426 tacgagctgcgagtacgctgagttcaccttgcacgagcaagcatgctgagtcgagtcgtgt 485
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     343 TACGACTCGCCAGTACGCTCGCTTCACCCTTGCACAGCAACATGCGTCGCGACGCGT 402

Qy     486 gctcgaagtgtgtgagcgcgcgcgcgcgcgcgcgctgtgtgtgagctgtgcgctcgtgtgc 545
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     403 ACTCAAGTCTCTAAG-AAAGAAGAA-G-CTCGTCTCTTAAGAGTCCTTAGA 462

Qy     546 tgtgagctcgttgcgagcgagcgctttgtgagcgtgtgtgtgagtgcgtgcgtgtgtgcgcgt 605
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     463 TGTGAGCTCGCCAGCGAGCGCTTTGTGAGCGTGTGTGTGAGTGCGTGCGTGTGTGCGCGT 522

Qy     606 cgtgtgcgtgcgtgcgtgcgtgcgtgcgtgcgtgcgtgcgtgcgtgcgtgcgtgcgtgcgtgc 665
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     523 CGTGTGCGTGCGTGCGTGCGTGCGTGCGTGCGTGCGTGCGTGCGTGCGTGCGTGCGTGC 582

Qy     666 gtgtgcgtgcgtgcgtgcgtgcgtgcgtgcgtgcgtgcgtgcgtgcgtgcgtgcgtgcgtgc 725
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     583 GTGTGCGTGCGTGCGTGCGTGCGTGCGTGCGTGCGTGCGTGCGTGCGTGCGTGCGTGC 642

Qy     726 gtgtgcgtgcgtgcgtgcgtgcgtgcgtgcgtgcgtgcgtgcgtgcgtgcgtgcgtgcgtgc 785
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     643 GTGTGCGTGCGTGCGTGCGTGCGTGCGTGCGTGCGTGCGTGCGTGCGTGCGTGCGTGC 702

Qy     786 gtgtgcgtgcgtgcgtgcgtgcgtgcgtgcgtgcgtgcgtgcgtgcgtgcgtgcgtgcgtgc 845
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     703 GTGTGCGTGCGTGCGTGCGTGCGTGCGTGCGTGCGTGCGTGCGTGCGTGCGTGCGTGC 762

Qy     846 gtgtgcgtgcgtgc 859
           ||||||||||||||
Db     763 GTGTGCGTGCGTGC 776


RESULT  10
BB478282
LOCUS       BB478282                 [illegible] bp    mRNA    linear   EST 25-OCT-2001
DEFINITION  BB478282 RIKEN full-length enriched, 13 days embryo heart Mus
```

Tagami,M., Taqawa,A., Fukahashi,F., Ikeda,Y., Tanaka,T., Toya,I., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
JOURNAL Unpublished (2001)
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45 503 9222
Fax: 81 45 503 9216
Email: genome-jas@gsc.riken.go.jp
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of CAP-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. , 10 (10), 1617-1630 (2000)
Waqi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,I., Matsuura ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. , 10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara ,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. , 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,E., Kiyosawa,H., Yamanaka,I., Aizawa ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y.
Computational Analysis of Full Length Mouse cDNAs Compared with Human Genome Sequences. Mamm. Genome, 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.
e mouse tissues.

FEATURES Location/Qualifiers
source 1..622
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="A330077F03"
/clone_lib "RIKEN full length enriched, adult male spinal cord"
/sex="male"
/tissue_type="spinal cord"
/dev_stage="adult"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGACAGAGAAGAGATCCAAGAGCTCGIIIIIIIIIIIIIIIVN 3']. cDNA was prepared by using thermoactivated reverse transcriptase and subsequently enriched for full length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 45.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAAATCCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC 1. "

BASE COUNT 151 a 157 c 172 g 142 t
ORIGIN

Query Match 5.8%; Score 293.6; DB 9; Length 622;
Best Local Similarity 69.1%; Pred. No. 6.5e-62;
Matches 401; Conservative 0; Mismatches 179; Indels 0; Gaps 0;

```
Qy      186 ccagacctcaagaaacactcgttcagatgaagacaaagatgqcaactgggatgcttqqqq 245
            || ||  || || |   | || ||||| ||||||   || |||  |||||||||| |||||
Db       25 CCTGAGTTCCAGGACTGCACGCTCAGAGGAAGACCGTGAAGGCCTCTGGGATGCCTGGGG 84

Qy      246 cgactgaagtgactgctcccgaacctgtqqqqaqaqaqcatcatattctctqcqaaqatq 305
            |   ||||| || || ||| | |||||||| || || || || ||||| ||| |  | ||
Db       85 CCCTTGGAGCGAGTGTTCCAGAACCTGTGGTGGGGGTGCCTCCTATTCCCTGAGACGCTG 144

Qy      306 tttgactgqaaqqaattgtgaaqqqcaqaacattcgqtacaaaacatqcaqcaatcatqa 365
            |||   | | ||  |||||||||   || ||||| || |  |||||||| |||   ||
Db      145 CCTGAGCAGCAAGAGCTGTGAAGGCAGAAATATTCCATATAGAACATGCAGTAATGTGGA 204

Qy      366 ctqccctccaqatqcaqaaqatttcaqaqcccaqcaqtqctcaqcctacaatqatqtccA 425
            |||||| ||||| |||   |||||| |||| || |||| || ||  |  || ||   |
Db      205 CTGCCCACCAGAAGCAGGTGATTTCCGAGCTCAACAGTGTTCTGCTCATAACGACGTCAA 264

Qy      426 gtatcaggqqqcattactatcaatqqctttcaqatatcaatqatcctqctqcccctqtqc 485
            ||| || || ||    ||||||||||||||   || |||||| || |    || || |
Db      265 GTACCATCGACAGCTTTATGAATGGCTTCCTGTATCTAATGACCCCGACAATCCATGCTC 324

Qy      486 actcaaqtqtcatqcacaaqqacaaaaacttqqtqqtqqaqctqqcactaaqqtactqqa 545
            |||||||||| || ||| ||||    | | |||| |||| || || ||| ||||  | ||
Db      325 ACTCAAGTGCCAAGCAAAACGAACCAGCCTGGTTGTGGAACTAGCACCGAAGGTCTTACA 384

Qy      546 tggaactcgttgcaacacggactccttqqacatgtqtatcaqtqqcatctqtcaqqcaqt 605
            ||| || || ||| | || || ||  |||| |||| |||| ||| | || ||    ||
Db      385 TGGTACGCGATGCTATACAGAATCACTGGATATGTGCATCAGCGGCCTATGCCAAATTGT 444

Qy      606 gqqctgcgatcggcaactgqqaaqcaatgccaaggaqqgacaactqtqqaqtctqtqccqq 665
            |||||| ||||  || |||| |||| || ||| |||||  ||||||||||| ||   |||
Db      445 TGGCTGTGATCACCAGCTGGGGAGCACTGTCAAAGAGGATAACTGTGGAGTGTGCCACGG 504

Qy      666 cqatqqctccacctqcaqqcttqtacqqqacaatcaaaqtcacacqtttctcctqaaaa 725
            |||||||| || ||| |||| || || || || |  || || || | ||| |     ||
Db      505 AGATGGCTCAACGTGCCGGCTCGTCCGAGGGCAGTATAAATCTCAGCTCTCTGCAAGTAA 564

Qy      726 aagagaagaaaaatgtaattgctgttcctttqqqaaqtcqa 765
            |   || || |  ||  | ||  |||||  || || ||
Db      565 ATCGGATGACACCGTGGTCGCCATTCCTTACGGCAGCAGA 604
```

RESULT 8
AQ743958
LOCUS AQ743958 890 bp DNA linear GSS 16-JUL-1999
DEFINITION HS_5504_A1_D01_T7A RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate=1080 Col=1 Row=G, DNA sequence.
ACCESSION AQ743958
VERSION AQ743958.1 GI:5521480
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 890)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)

```
            or from Research Genetics (info@resgen.com). BAC end Web Server:
            http://www.hgsc.washington.edu
            Plate: 1080  row: G  column: 1
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 800.
FEATURES             Location/Qualifiers
     source          1..890
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate 1080 Col 1 RowG"
                     /clone_lib "RPCI-11 Human Male BAC Library"
                     /sex "male"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     Male blood DNA was isolated from one randomly chosen donor
                     and partially digested with a combination of EcoRI and
                     EcoRI Methylase. Size selected DNA was cloned into the
                     pBACe3.6 vector at EcoRI sites"
BASE COUNT       228 a    185 c    218 g    259 t
ORIGIN


  Query Match             25.6%;  Score 281.8;  DB 12;  Length 890;
  Best Local Similarity   97.6%;  Pred. No. 1.1e-64;
  Matches  286;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy    1696 ccaaacgttcattccagaacc tgatcagcctcgtcgattacaacgtatgagccaagtgtcgaa 1755
           || | ||||||||||||||||a||||||||||||||||||-|||||||||| ||||||||
Db     125 CCTAGGTTCATTCCAGAACCCTGGTCAGCCTGCAGTACCACGTGTGGGCCGGGTGTGCAG 184

Qy    1756 gtccgcgagatgaagtgccgtgtgctctcgacattacgcgaagctgatgactgagtacc 1815
           |||| |||||||| ||||||||||||||||||||||||||||||||| | ||||||||
Db     185 GTCCGTGAGGTGAAGTGCCGTGTGCTCTCACATTCACGCAGACTGATATGACGCCG 244

Qy    1816 gagggaagatgtgtaaagccccaagtgccaccaacgcccgtgccttcgaaaacgat 1875
           |||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db     245 GAGGAAGACTGTGAAGGCCCAAGTGCCCACCGAACGGCCCTGCCTCCTGGAAGCATGT 304

Qy    1876 gatgatgagccccgcctccccgaaagctagacatccctctccctgaggacagtgagacgact 1935
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     305 GATGAGAGCCCGGCTCCCGAGAGCTAGACATCCCTCTCCCTGAGGACAGTGAGACGACT 364

Qy    1936 tacgactgatatgtacgtgatttacccctgcacagcaacatgcttgaggag 1988
           ||||||||||C||||||||||||||||||||||||||||||||||| ||||||
Db     365 TACGACTGGGAGTACGCTGGGTTCACCCCTTGCACAGCAACATGCGTGGGAGG 417


RESULT   9
BE888902
LOCUS       BE888902                 850 bp    mRNA    linear   EST 25-OCT-2000
DEFINITION  601513928F1 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:4916294 5',
            mRNA sequence.
ACCESSION   BE888902
VERSION     BE888902.1  GI:8456?0
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 850)
  AUTHORS   NIH_MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health. Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-rmail@nih.gov
            Tissue Procurement: ASTP
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9238  row: a  column: 07
            High quality sequence stop: 742.
FEATURES             Location/Qualifiers
     source          1..850
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4916294"
                     /clone_lib "NIH_MGC_71"
                     /tissue_type "leiomyosarcoma"
                     /lab_host "DH10B (phage-resistant)"
                     /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: oligo dT.
                     Average insert size 2.1 kb.  "
BASE COUNT       202 a    213 c    228 g    207 t
ORIGIN


  Query Match             25.4%;  Score 278;  DB 12;  Length 850;
  Best Local Similarity   97.9%;  Pred. No. 1.6e-64;
  Matches  451;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy     186 ccagaccttcaatgcagaaccctggtcagcctgcagtaccacgtgtgggccgggtgtgcag 245
           || ||  || || | | | | ||  | | || |||  || |||  ||||||| ||||| |
Db     120 CCTAGGTTCATTCCAGAACCCTGGTCAGCCTGCAGTACCACGTGTGGGCCGGGTGTGCAG 179

Qy     246 gtccgcgagatgaagtgccgtgtgctctcgacattacgcgaagctgatgactgagtacc 305
           | | |||||| |||| | | |||| || || || || ||||| || | ||
Db     180 GTCCGTGAGGTGAAGTGCCGTGTGCTCTCACATTCACGCAGACTGATATGACGCCG 237

Qy     306 tttgactgtaaagtaattgtgaaagtaaaacattcagtaccaagacatgcagcaatcatga 365
           |||  | | ||  ||||||||  || || || ||||  ||||||| |||  ||
Db     238 GTGAGGAAGTGAAGTGCCGTGTGCTCTCACATTCACGCAGACTGATATGACGCCG 295

Qy     366 ttcgccccaaatgccgtcatattcctaagcccaagcctgctgcctcaacaatgatgtccca 425
           | ||| | |||| ||| | | ||| | | ||||| | ||||| | ||||||||| |
Db     296 GTGAGGAAGTGAAGTGCCGTGTGCTCTCACATTCACGCAGACTGATATGACGCCG 353

Qy     426 gtatcagggagcattactatgaatagttccagcgatatgaatgatctgtctgtcccccatatgc 485
           | | || || || |  ||||||||||||||    | ||||| |||| |||  ||| ||| |
Db     354 GTACCATGGCCAGTTTTATGAATGCTTCCTGTGTCTAATGACCCTGACAACCCATGTTC 412

Qy     486 actcaatgatgtcatgcaaaaaacaaaactgaatagtgtgaagctgacgacctaaagtactgga 545
           |||||||| || ||  ||||  || | |||| || || || ||||||||||  | ||
Db     413 ATTCAAGTGTCCAAGTCAAAAAAAA AACCTGTTGTTGAACTAGCACTAAGGTCTTAGA 472

Qy     546 tgaaactgttgcaagcagaagctcttgaacatgtgtatgaagtgcatctgtcagacagt 605
           | | | || || | | | | | ||||| |||| ||||||||  | || ||    ||
Db     473 AGCTACGCG TGCTATACAGAATCTTGGATATGTGCATCAGTGGTTTATGCCAAATTGT 532

Qy     606 agactgcgatgcggcaactgaaaacaatgccaaagaagagcaactgtgaagtctgtgcgag 665
           | |||||||| || ||||| | | |||||| || ||||||| |||||   |||
Db     533 TTTTTTATATCATAAAAAAAAACAGTGAAGAAATAAAGGGGGGTCTGCAACGG 581

Qy     666 caatagatcaccatgcagacttatgaaggacaatcaaatcacacattttcactgaa 723
           | | || |||||||  |||| | || || || | || || || | || |
Db     582 AGATGGGGTCACATCTTTTGTTGAGGAGTATAAATCCCACTTCAGCAACCC 641

Qy     724 aaaagtaaaaaaaatatgattgctgttctttgtgaaatgcgaaatgtgaagaattacagt 782
           ||  | || || | ||  |||   |||| | || | |  | | |||    || |
Db     642 AATGGGATGATACTTTTTTCAAATTCCCTATGGCAAGTTGACTATTTCGCTTGGTTT 701

Qy     783 aaaaaacctaaa aaaaaaaaatgaatcaaaaacacttcaaaaaaatcaaaaaaaaaca 842
           | | |||| | |  |  |  | | ||||||      ||| ||| | | |
Db     702 AAAAGCCCTGATGACTCTCTATCAGAACAAAACCTCCACGGGACTCAAGGTGGCAAC 761

Qy     843 caaattttaac 853
           | | || |
Db     762 AGGGCTTATC 771


RESULT   10
BB478282
LOCUS       BB478282                 815 bp    mRNA    linear   EST 25-OCT-2001
DEFINITION  BB478282 RIKEN full-length enriched, 13 days embryo heart Mus
```

Set to version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 19, 2002, 21:55:25 ; Search time 0.08 Seconds
(without alignments)
13732.264 Million cell updates/sec

Title:
Perfect score: 5076
Sequence: 1 atgattctcgatcaaaaaa.....agtcatgtcatcaaaaat 7218

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 884544 seqs, 1208165752 residues

Total number of hits satisfying chosen parameters: 884544

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ispat/US09_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/ispat/US09_PUBCOMB.seq:*
3: /cgn2_6/ptodata/1/ispat/US09_PUBCOMB.seq:*
4: /cgn2_6/ptodata/1/ispat/US09_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/ispat/US09_PUBCOMB.seq:*
6: /cgn2_6/ptodata/1/pubpatfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result. The better matches are printed, and it is derived by analysis of the data score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	5076	100.0	7218	2	US-08-232-463-14	Sequence 14, Appl
2	4156	81.9	2051	2	US-08-672-814D-1	Sequence 1, Appli
3	4156	81.9	2051	4	US-09-333-696-1	Sequence 1, Appli
4	3962	78.1	1426	4	US-08-985-526-4	Sequence 4, Appli
5	3958	78.0	1470	4	US-09-099-041A-27	Sequence 27, Appl
6	3958	78.0	1563	4	Patent No. 5618685	Patent No. 5618685
7	3958	78.0	2050	1	US-09-099-041A-9	Sequence 9, Appli
8	3958	78.0	3080	1	US-09-099-041A-25	Sequence 25, Appl
9	3958	78.0	3386	4	US-09-099-041A-7	Sequence 7, Appli
10	3944	77.7	748	4	Patent No. 5169845 14	Patent No. 5169845
11	3941	77.7	2054	4	US-08-468-859B-8	Sequence 8, Appli
12	3944	77.7	2059	4	US-08-468-859A-8	Sequence 8, Appli
13	3942	77.7	1573	4	Patent No. 5169845 5	Patent No. 5169845
14	3942	77.7	1591	4	US-08-468-859B-9	Sequence 9, Appli
15	3942	77.7	1591	4	US-08-468-859A-9	Sequence 9, Appli
16	3942	77.7	1457	4	Patent No. 5169845 11	Patent No. 5169845
17	3942	77.7	2376	1	US-07-912-952-8	Sequence 8, Appli
18	3942	77.7	4410365	1	US-09-163-840A-2	Sequence 2, Appli
19	3942	77.7	4411529	1	US-09-163-840A-1	Sequence 1, Appli
20	3944	77.7	4605	2	US-08-231-193A-52	Sequence 52, Appl
21	3944	77.7	4605	2	US-08-486-274A-52	Sequence 52, Appl
22	3944	77.7	4605	4	US-08-940-095A-52	Sequence 52, Appl
23	3944	77.7	4605	4	US-08-940-095A-52	Sequence 52, Appl
24	3936	77.5	445	4	Patent No. 5169845 2	Patent No. 5169845
25	3936	77.5	2958	4	Patent No. 5169845 1	Patent No. 5169845
26	3936	77.5	1104	4	US-08-869-908-16	Sequence 16, Appl
27	3928	77.4	1164	2	US-08-445-640-5	Sequence 5, Appli
28	3928	77.4	1164	4	US-08-170-558-5	Sequence 5, Appli
29	3928	77.4	1164	4	US-08-447-814-5	Sequence 5, Appli
30	3928	77.4	1164	4	US-08-445-401-5	Sequence 5, Appli
31	3928	77.4	1807	4	Patent No. 5510466 1	Patent No. 5510466
32	3928	77.4	2742	1	US-08-209-x-5A-2	Sequence 2, Appli
33	3928	77.4	2742	2	US-08-441-104A-2	Sequence 2, Appli
34	3928	77.4	2742	2	US-08-443-816A-2	Sequence 2, Appli
35	3928	77.4	2742	4	US-08-417-981A-2	Sequence 2, Appli
36	3928	77.4	3611	1	US-08-445-640-1	Sequence 1, Appli
37	3928	77.4	3611	4	US-08-170-558-1	Sequence 1, Appli
38	3928	77.4	3611	4	US-08-447-814-1	Sequence 1, Appli
39	3928	77.4	3611	4	US-08-445-401-1	Sequence 1, Appli
40	3926	77.3	4843	4	US-08-986-485-2	Sequence 1, Appli
41	3926	77.3	9562	4	US-08-886-987-1	Sequence 1, Appli
42	3926	77.3	9562	4	US-09-306-949-1	Sequence 1, Appli
43	3924	77.3	551	4	US-08-750-141A-5	Sequence 5, Appli
44	3924	77.3	2313	1	US-08-232-538-5	Sequence 5, Appli
45	3924	77.3	2362	1	US-08-232-538-17	Sequence 17, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5677367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: WOLKERSTORFER, H.
; TITLE OF INVENTION: RECOMBINANT HEMAGGLUTININ VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22314-2344
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/995,613
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.8
; FILING DATE: 29-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENI, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/134 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pIZup1 #1s
US-08-232-463-14

```
  Query Match             1.1%;  Score 56;  DB 1;  Length 7218;
  Best Local Similarity   9.2%;  Pred. No. 1.5e-06;
  Matches   41;  Conservative  215,  Mismatches  190,  Indels    0,  Gaps    0;

Qy    4036 aaqqaaatatatttqaaatqaaaccacaaaqaaaqcaaatacttttqqaaatcaacatqqcacaa 4095
           ||  ||| :  ||:| : ::  : :: ::::  :: ::   :::::  : : :: : ::
Db    1452 CATACAACAAT!TC!!A'RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1294

Qy    4096 aatqaqqcaaatqttqaaataaccaaaaatqacctttatctqqatqatqaccacattaqtaa 4155
           :: :::: ::: : :::: :: ::::: ::     :   ::: :: ::  : :  :: ::
Db    1392 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1333

Qy    4156 ccaqcctttctttaaaqctctqttaqqccactqcaqcaattttccaqqaaqccaccaactc 4215
             ::       :::::   :  :::  :  : :: ::    : :::::: :  ::
Db    1332 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1273

Qy    4216 ctttttttgtgtaataaa gttttttg atgatttaaaaaaaaaatatatattta 4277
             :::::  :::::: ::: ::   :::: :: ::  ::: :::::: : : :: : :::
Db    1272 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1213

Qy    4278 tacaqcaaaqttqatqaqttataaaaaacatqaatcaqctatqaaaccqaaaaaat 4335
           : ::   ::   :: ::: :: :::::: : :::  ::   : :::::  :::::::
Db    1212 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1153

Qy    4336 caqcqatqatcttgcqtcaaaqctqatatatcaqctqtqaqcqaattaqccaaaggcaca 4395
           :: :: ::   : :   :: :: :: :: :: :: ::  :: :: :: :: :: :: : :
Db    1152 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1093

Qy    4396 gcaaaaaaatatqaatqqtqc qtttcat cattgatqqq c t t t t t qqcc at 4455
           :  :: : : : : :: :: :: :  :   :   : :  !!!!! !!  !!!!! :
Db    1092 RRRRRRRRRRRRRRRRRRRRRRRRRRRAIGCAAAGCTCCTTCGAGCTGCAAGCCAAAGCTCGG 1033

Qy    4456 aqqqqaaacaqqqaqtqtqcccaaa 4481
           :   :  : : ::: :: :  ::::
Db    1032 AATTAATTCTGTGAGGCTATGGCAAA 1007


RESULT   2
US-08-672-814D-1/c
; Sequence 1, Application US/08672814D
; Patent No. 5952480
;  GENERAL INFORMATION:
;    APPLICANT:  Leung, David W.
;    APPLICANT:  Weeks, Keitha
;    TITLE OF INVENTION:  MAMMALIAN
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cell Therapeutics, Inc.
;      STREET:  201 Elliott Avenue West
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  U.S.A.
;      ZIP:  98119
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" disk, 1.44Mb, double side, high density
;      COMPUTER:  PC Clone (486 microprocessor)
;      OPERATING SYSTEM:  MS DOS Version 6.1, Windows NT
;      SOFTWARE:  WORD 6.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/672,814
;      FILING DATE:  28-Jun-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Paciszewski, Stephen
;      REGISTRATION NUMBER:  36,131
;      REFERENCE/DOCKET NUMBER:  1867
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206)282-7100
;      TELEFAX:  (206)284-6206
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2051
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double stranded
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    HYPOTHETICAL:  no
;    ANTI-SENSE:  no
;    FRAGMENT TYPE:
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapien
;      STRAIN:
;      INDIVIDUAL ISOLATE:
;      DEVELOPMENTAL STAGE:
;      HAPLOTYPE:
;      TISSUE TYPE:
;      CELL TYPE:
;      CELL LINE:
;      ORGANELLE:
;    FEATURE:
;      NAME/KEY:  CDS
US-08-672-814D-1


  Query Match             5.6%;  Score 41.6;  DB 2;  Length 2051;
  Best Local Similarity   55.6%;  Pred. No. 0.022;
  Matches   80;  Conservative    6;  Mismatches   64;  Indels    0;  Gaps    0;

Qy     599 aqqcaqtqqactqcqatcggcaactaqqaaaqcaatqccaaaqaaqacaactqtqqaqtct 658
           | | | ||| | |     || | || |||||||     | ||||||||
Db    1060 AAGTACTGGTATTTGGATAACACATAGGCAGCAATGAATCCAAACACAACTGTGGAAAAG 1001

Qy     659 qtqcqcqcatqqctcaqctcaaqaqctttatqcqqqaqcaatcaaaqtccaqcqttctc 718
           ||  | |||  |||  |   || ||| |  || |||||   || | |    |||||
Db    1000 AAACCACCAATGAATCCTTCCCAACTCTTTTATGCAGACAACTTAATTAATGGAGTTCTC 941

Qy     719 ctqaaaaaaqaqaaqaaaatqtaa 742
           |  |||||| |    |||| ||||
Db     940 CCAAAAAAAAAATCCAAAAAGGTAA 917


RESULT   3
US-09-333-696-1/c
; Sequence 1, Application US/09333696
; Patent No. 6200769
;  GENERAL INFORMATION:
;    APPLICANT:  Leung, David W.
;    APPLICANT:  Weeks, Keitha
;    TITLE OF INVENTION:  MAMMALIAN
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cell Therapeutics, Inc.
;      STREET:  201 Elliott Avenue West
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  U.S.A.
;      ZIP:  98119
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" disk, 1.44Mb, double side, high density
;      COMPUTER:  PC Clone (486 microprocessor)
;      OPERATING SYSTEM:  MS-DOS Version 6.1, Windows NT
;      SOFTWARE:  WORD 6.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/333,696
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/672,814
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Paciszewski, Stephen
;      REGISTRATION NUMBER:  36,131
;      REFERENCE/DOCKET NUMBER:  1867
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206)282-7100
;      TELEFAX:  (206)284-6206
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
```

```
;       LENGTH:  2951
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double stranded
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA to mRNA
;     HYPOTHETICAL:  no
;     ANTI-SENSE:  no
;     FRAGMENT TYPE:
;     ORIGINAL SOURCE:
;       ORGANISM:  Homo sapiens
;       STRAIN:
;       INDIVIDUAL ISOLATE:
;       DEVELOPMENTAL STAGE:
;       HAPLOTYPE:
;       TISSUE TYPE:
;       CELL TYPE:
;       CELL LINE:
;       ORGANELLE:
;     FEATURE:
;       NAME/KEY:  CDS
US-09-333-606-1

  Query Match                 71.0%;  Score 41.6;  DB 4;  Length 2951;
  Best Local Similarity   55.6%;  Pred. No. 0.022;
  Matches   80;  Conservative    0;  Mismatches   64;  Indels    2;  Gaps    2;

Qy     599 aggcaatgaggctgcgatgcgggcctcgaggcagtgccagaggaggacacgtgtgaggtgt 658
           |||| ||| | | |     || | |||||||      | ||||||||
Db    1060 AAGTACTGGTATTGGATAACACATAGGCAGTAATGAATCCAAACACAACTGTGGAAAAG 1001

Qy     659 atgccttcctatgctctcctctcctttctatcgggaggcatcaagtccgtttttcc 715
           || | | ||  ||| |  || ||| |  || ||||   || || |     ||||
Db    1000 AAACCAGCAATGAATCTTCCCAATCTTTTTAGGAGACAACTTAATTAATCGATTTTCT 941

Qy     716 ctcgaaaaaaaaaaaaaaaaaaatgcct 742
           | ||||||| |    |||| ||||
Db     940 CCAAAAAAAAATCCAAAAAGTTAA 917


RESULT  4
US-08-985-526-4
; Sequence 4, Application US/08985526
; Patent No. 6080720
; GENERAL INFORMATION:
;     APPLICANT:  Mixson, James A
;     TITLE OF INVENTION:  PARTICLE-DNA COMPLEXES CONTAINING DNA
;     TITLE OF INVENTION:  ENZYMES AND ANTISENSE OLIGOMERS AND USES THEREOF IN GENE
;     TITLE OF INVENTION:  THERAPY
;     NUMBER OF SEQUENCES:  43
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Connolly, Bove, Lodge, & Hutz
;       STREET:  1220 Market Street, P.O. Box 2207
;       CITY:  Wilmington
;       STATE:  Delaware
;       COUNTRY:  U.S.A.
;       ZIP:  19899
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC Compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/985,526
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/298,041
;       FILING DATE:  26-AUG-1994
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  McMorrow Jr., Robert J.
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (302) 658-9141
;       TELEFAX:  (302) 658-5614
;     INFORMATION FOR SEQ ID NO:  4:
;       SEQUENCE CHARACTERISTICS:
;         LENGTH:  1426 base pairs
;         TYPE:  nucleic acid
;         STRANDEDNESS:  single
;         TOPOLOGY:  linear
US-08-985-526-4

  Query Match                 0.7%;  Score 40.2;  DB 4;  Length 1426;
  Best Local Similarity   60.8%;  Pred. No. 0.04;
  Matches   59;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy    1276 cctgaatgtgcatattccgtgtctctgtgaggaggaatttcaaaacaaaaggtttgtgtat 1335
           || |||       ||||| || ||||||||||||||| | |||| ||| ||   | | ||
Db     598 CCATGGGACATCTGCTCTGTACCTGTCGTACGGAGGGTTACAAAACTTATGTCTCTCTAT 657

Qy    1336 gtaatgaggaatccatgcatgtaaaaaaatattgcaggtgt 1375
           ||  ||       ||| || || | |      || ||
Db     658 GTCGACTCTAGAATGACTGAAGACAACAAAGACTTGCGT 696


RESULT  5
US-09-099-041A-27
; Sequence 27, Application US/09099041A
; Patent No. 6440576
; GENERAL INFORMATION:
;  APPLICANT: Bertin, John
;  TITLE OF INVENTION: NOVEL MOLECULES OF THE CARD RELATED
;  TITLE OF INVENTION: PROTEIN FAMILY AND USES THEREOF
;  FILE REFERENCE: 07364.076001
;  CURRENT APPLICATION NUMBER: US/09/099,041A
;  CURRENT FILING DATE:  1998-06-17
;  PRIOR APPLICATION NUMBER: 60/049,942
;  PRIOR FILING DATE: 1998-02-06
;  NUMBER OF SEQ ID NOS: 37
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 27
;   LENGTH: 1470
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-099-041A-27

  Query Match                 0.7%;  Score 35.8;  DB 4;  Length 1470;
  Best Local Similarity   49.2%;  Pred. No. 1.2;
  Matches   61;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy    2109 tcccccaggtgaccatgtgagtctgtgaaggcccctgtcaatcatgtgtgtgaggtgat 2168
           |||||| | | | | |  | | | |||| | |||| |  | | ||||
Db     707 tcccccacgtggccctctcacctttcgatggccctgacagtagtgcacacctgcactcacc 766

Qy    2169 ttaaacccccgatgtatgtatgctgtcccccccccccccccctgtgtctctgtgtgtgtgtgt 2228
           || | | | |  |    |||| || || |  ||| |||| | |  | |  |
Db     767 tgagcgctctgactgagactctgcccctgagtgtgccccctgtgtctctgtgtgtgtgtgt 826

Qy    2229 ccccgtgtgagagccccatgctttacaaatcatccatgtatttgtgtgccctcctgtgtgtgt 2288
           | | |  ||  ||   | | |  |||  ||  ||  |  |  |  ||  | |
Db     827 ccccacgtgcgtgtgtgtgagcgtgtgcgtgagcgtgcgtgtgtgtgtgtgtgtgtgtgt 886

Qy    2289 gtca 2292
           ||||
Db     887 gtca 890


RESULT  6
S16983  c
;Patent No. 5164483
;  APPLICANT: WAI YEE, CHAN
;  TITLE OF INVENTION: PREGNANCY SPECIFIC PROTEINS APPLICATIONS
;  NUMBER OF SEQUENCES: 48
;  CURRENT APPLICATION DATA:
```

```
;       APPLICATION NUMBER:  US/07/395,433
;       FILING DATE: 07 AUG 1989
;SEQ ID NO:8
;       LENGTH: 1694
5169a45-8


  Query Match             0.7%;  Score 35.8;  DB 6;  Length 1694;
  Best Local Similarity   65.8%;  Pred. No. 1.8;
  Matches   52;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy   4021 aatcaacaatcatttcctttgcttttcaatcgcatcccctgttgttcacaaatgttttcccttcaaa 4080
           |||| | |  |   |  |   || ||| | |||||||  | | | |||||||  |||   |
Db    142 aatcaaaaaaaaaaaaatatattccaacgcatccctctgctgatcaaaatgtccaccccgaaag 201

Qy   4081 aatcaaaaaaacctcagtct 4099
           ||| ||||||||||  ||
Db    202 aatcaaaaaaactaaaatt 220


RESULT   7
US-09-099-041A-9
; Sequence 9, Application US/09099041A
; Patent No. 6340576
; GENERAL INFORMATION:
;  APPLICANT: Bertin, John
;  TITLE OF INVENTION: NOVEL MOLECULES OF THE CARD-RELATED
;  TITLE OF INVENTION:  PROTEIN FAMILY AND USES THEREOF
;  FILE REFERENCE: 07334-076001
;  CURRENT APPLICATION NUMBER: US/09/099,041A
;  CURRENT FILING DATE:  1998-06-17
;  PRIOR APPLICATION NUMBER: 09/019,942
;  PRIOR FILING DATE: 1998-02-06
;  NUMBER OF SEQ ID NOS: 37
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9
;   LENGTH: 2859
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-099-041A-9


  Query Match             0.7%;  Score 35.8;  DB 4;  Length 2859;
  Best Local Similarity   49.7%;  Pred. No. 1.9;
  Matches   91;  Conservative    0;  Mismatches   92;  Indels    0;  Gaps    0;

Qy   2109 tccccccaaatgaaatgtaggctcttgaggagcctgctcaactacctatgaaattgaaat 2168
           ||||||  |  | | | |  |  | |  |||| |  |||| |  ||| |||||
Db    824 tccccacatgaccctttcacttcaatgcaggctgtcaacaactacctatgaattgaaac 883

Qy   2169 tcaggcccaaaatgtgtactgctgcacaccaaggaggaagcccttcccctcctaaaaata 2228
           | || | || |     |  |||| ||  || |   |||  |||| |  |   |
Db    884 tgagccctgctgctgtcaactctgccctgaggagctgccaaccccctgatcttgctga 943

Qy   2229 ccaaaatgaaaaggccatatcttcaaaatgcaatcatttgaatcaattcaatctgctcag 2288
           || |  ||   ||   | |||  ||||  || | ||  ||   |  ||  | |
Db    944 ccaacctgctcaatgaaaagctgctcaaggaagctgctcaacagctgctcaacag 1003

Qy   2289 gca 2291
           |||
Db   1004 gca 1006


RESULT   8
US-09-099-041A-25
; Sequence 25, Application US/09099041A
; Patent No. 6340576
; GENERAL INFORMATION:
;  APPLICANT: Bertin, John
;  TITLE OF INVENTION: NOVEL MOLECULES OF THE CARD-RELATED
;  TITLE OF INVENTION:  PROTEIN FAMILY AND USES THEREOF
;  FILE REFERENCE: 07334-076001
;  CURRENT APPLICATION NUMBER: US/09/099,041A
;  CURRENT FILING DATE:  1998-06-17
;  PRIOR APPLICATION NUMBER: 09/019,942
;  PRIOR FILING DATE: 1998-02-06
;  NUMBER OF SEQ ID NOS: 37
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 25
;   LENGTH: 3080
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(1470)
US-09-099-041A-25


  Query Match             0.7%;  Score 35.8;  DB 4;  Length 3080;
  Best Local Similarity   49.7%;  Pred. No. 2;
  Matches   91;  Conservative    0;  Mismatches   92;  Indels    0;  Gaps    0;

Qy   2109 tccccccaaatgaaatgtaggctcttgaggagcctgctcaactacctatgaattgaaat 2168
           ||||||  |  | | | |  |  | |  |||| |  |||| |  ||| |||||
Db    767 tccccacatgaccctttcacttcaatgcaggctgtcaacaactacctatgaattgaaac 826

Qy   2169 tcaggcccaaaatgtgtactgctgcacaccaaggaggaagcccttcccctcctaaaaata 2228
           | || | || |     |  |||| ||  || |   |||  |||| |  |   |
Db    827 tgagccctgctgctgtcaactctgccctgaggagctgccaaccccctgatcttgctga 886

Qy   2229 ccaaaatgaaaaggccatatcttcaaaatgcaatcatttgaatcaattcaatctgctcag 2288
           || |  ||   ||   | |||  ||||  || | ||  ||   |  ||  | |
Db    887 ccaacctgctcaatgaaaagctgctcaaggaagctgctcaacagctgctcaacag 946

Qy   2289 gca 2291
           |||
Db    947 gca 949


RESULT   9
US-09-099-041A-7
; Sequence 7, Application US/09099041A
; Patent No. 6340576
; GENERAL INFORMATION:
;  APPLICANT: Bertin, John
;  TITLE OF INVENTION: NOVEL MOLECULES OF THE CARD-RELATED
;  TITLE OF INVENTION:  PROTEIN FAMILY AND USES THEREOF
;  FILE REFERENCE: 07334-076001
;  CURRENT APPLICATION NUMBER: US/09/099,041A
;  CURRENT FILING DATE:  1998-06-17
;  PRIOR APPLICATION NUMBER: 09/019,942
;  PRIOR FILING DATE: 1998-02-06
;  NUMBER OF SEQ ID NOS: 37
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 7
;   LENGTH: 3382
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (245)...(3103)
US-09-099-041A-7


  Query Match             0.7%;  Score 35.8;  DB 4;  Length 3382;
  Best Local Similarity   49.7%;  Pred. No. 2.2;
  Matches   91;  Conservative    0;  Mismatches   92;  Indels    0;  Gaps    0;

Qy   2109 tccccccaaatgaaatgtaggctcttgaggagcctgctcaactacctatgaattgaaat 2168
           ||||||  |  | | | |  |  | |  |||| |  |||| |  ||| |||||
Db   1068 tccccacatgaccctttcacttcaatgcaggctgtcaacaactacctatgaattgaaac 1127

Qy   2169 tcaggcccaaaatgtgtactgctgcacaccaaggaggaagcccttcccctcctaaaaata 2228
           | || | || |     |  |||| ||  || |   |||  |||| |  |   |
Db   1128 tgagccctgctgctgtcaactctgccctgaggagctgccaaccccctgatcttgctga 1187
```

```
Qy    2221 cccctttgtgggcccctgtgtttcaaagcatgcaatcaattgcaacgacccctgacgtcc 2280
            ||| | || || |  | |||  |||| || | | ||  |  | |  | |
Db    1188 ccaacttgtctgtgtgaaattctttgaaaggtgttaacgttgcgtgcagggactgcagg 1247

Qy    2281 gcca 2284
            ||||
Db    1248 gcca 1251


RESULT 10
S16085 14
;Patent No. 5160965
;  APPLICANT: WAI YEE, CHAN
;  TITLE OF INVENTION: PREGNANCY SPECIFIC PROTEINS APPLICATIONS
;  NUMBER OF SEQUENCES: 48
;  CURRENT APPLICATION DATA:
;    APPLICATION NUMBER:  US/07/xxx,xxx
;    FILING DATE: 07 AUG 1989
;SEQ ID NO:14:
;    LENGTH: 2284
S16085 14


  Query Match                 0.2%;  Score 44.4;  DB 5;  Length 2284;
  Best Local Similarity   54.8%;  Pred. No. 2.1;
  Matches   68;  Conservative    0;  Mismatches   56;  Indels    2;  Gaps    2;

Qy    4021 aatgcgcatgattttcttgtctttttcaatgaatcccgtgttgtcgcaatattttcctttaaa 4080
           ||||| | |       +  | |||| ||||| || | | |||||||   | |  ||
Db     279 aatttaaaatttaaaatatatttcaatgatatccctgtatcaacaatatcttcgttttt 338

Qy    4081 aatgaaggaaatctgcgtctgcatgcgcaatgctctttaaaaaaaaatgcatgcgcatatt 4140
           ||| ||| |||||  |       || ||   |       |       |  |  |
Db     339 aatgcaggatatcgcgttacgtatgaaacgagtgaaaaaaaaatgcatgcgcatatt 398

Qy    4141 gtac 4144
           ||||
Db     399 gtac 402


RESULT 11
US/08/468,856B/8
; Sequence 8, Application US/08468856B
; Patent No. 5652122
  GENERAL INFORMATION:
;    APPLICANT:  Barnett, Thomas; Elliot, James; Kamarck, Michael;
;    APPLICANT:  Kretschmer, Axel
;    TITLE OF INVENTION:  DNAs CODING FOR MEMBERS OF THE
;    TITLE OF INVENTION:  CAPTIN EMBRYONIC ANTIGEN FAMILY
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Spencer Born Kramer & Woods
;      STREET:  660 White Plains Road
;      CITY:  Tarrytown
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10591-5144
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 5.25 inch, 2.0 Mb storage
;      COMPUTER:  APPLE MACINTOSH 6500
;      OPERATING SYSTEM:  SYSTEM 7.5
;      SOFTWARE:  WordPerfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/468,856B
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/207,678
;      FILING DATE:  16-JUN-1988
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/xxx,xxx
;      FILING DATE:  07-AUG-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/274,107
;      FILING DATE:  21-NOV-1988
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/207,678
;      FILING DATE:  16-JUN-1988
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/xxx,xxx
;      FILING DATE:  19-JUN-1987
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/019,xxx
;      FILING DATE:  19-FEB-1987
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/06/898,xxx
;      FILING DATE:  13-AUG-1986
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kurt G. Briscoe
;      REGISTRATION NUMBER:  33,141
;      REFERENCE/DOCKET NUMBER:  0051.212.10.K0B
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (914) 332-1700
;      TELEFAX:  (914) 332-1844
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2009 nucleotides
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US/08/468,856B/8


  Query Match                 0.2%;  Score 44.4;  DB 5;  Length 2009;
  Best Local Similarity   54.8%;  Pred. No. 4.2;
  Matches   68;  Conservative    0;  Mismatches   56;  Indels    2;  Gaps    2;

Qy    4021 aatgcgcatgattttcttgtctttttcaatgaatcccgtgttgtcgcaatattttcctttaaa 4080
           ||||| | |       |  || |||| |||||| | | |||||||   | |  ||
Db     370 aatttaaaatttaaaatatatttcaatgatatccctgtatcaacaatatcttcgttttt 429

Qy    4081 aatgaaggaaatctgcgtctgcatgcgcaatgctctttaaaaaaaaatgcatgcgcatatt 4140
           ||| |||| | | || |  | | |  | | |  |  |  |  | | | |
Db     430 gtaaaatgcaaatctacaacttaaatttaaaatacataaaaaaaaatgcatgcgcatatt 489

Qy    4141 gtac 4144
           ||||
Db     490 gtac 493


RESULT 12
US/08/468,859A/8
; Sequence 8, Application US/08468859A
; Patent No. 6022958
  GENERAL INFORMATION:
;    APPLICANT:  Barnett, Thomas; Elliot, James; Kamarck, Michael;
;    APPLICANT:  Kretschmer, Axel
;    TITLE OF INVENTION:  DNAs CODING FOR MEMBERS OF THE
;    TITLE OF INVENTION:  CAPTIN EMBRYONIC ANTIGEN FAMILY
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Spencer Born Kramer & Woods
;      STREET:  660 White Plains Road
;      CITY:  Tarrytown
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10591-5144
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 5.25 inch, 2.0 Mb storage
;      COMPUTER:  APPLE MACINTOSH 6500
;      OPERATING SYSTEM:  SYSTEM 7.5
;      SOFTWARE:  WordPerfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/468,859A
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  435
```

```
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/027,974
;       FILING DATE:  08-MAR-1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/760,031
;       FILING DATE:  13-SEP-1991
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/274,107
;       FILING DATE:  21-NOV-1988
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/[illegible]
;       FILING DATE:  16-JUN-1988
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/060,031
;       FILING DATE:  19-JUN-1987
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/016,683
;       FILING DATE:  19-FEB-1987
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 06/896,361
;       FILING DATE:  13-AUG-1986
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Kurt G. Briscoe
;       REGISTRATION NUMBER:  34,141
;       REFERENCE/DOCKET NUMBER:  MEI 242.9 KGB
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (914) 332-1700
;       TELEFAX:  (914) 332-1844
;   INFORMATION FOR SEQ ID NO:  8:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2009 nucleotides
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US-08-468-859A-8


  Query Match                 0.7%;  Score 34.4;  DB 4;  Length 2009;
  Best Local Similarity   54.8%;  Pred. No. 4.2;
  Matches   68;  Conservative    0;  Mismatches   56;  Indels    0;  Gaps    0;

Qy    4021 aatgcaatgtttcattgcttcaatgatcctgttgttgaaaatttccttgaa 4080
           ||||| | |       |  || |||| ||||||| || | ||||||| | ||  ||
Db     370 AGTGGAAGAGAAAGAGTATATTCCAATGCATCCCTGCTGATCCAGAATGTCACGCAGGAC 429

Qy    4081 aatgaaggaaactacgtct 4099
           ||| |||| ||||| ||
Db     388 gatgcagggatcctacact 406


RESULT  14
US-08-468-856B-9
; Patent No. 6013772
;  GENERAL INFORMATION:
;    APPLICANT:  Barnett, Thomas; Elting, James; Kamarck, Michael;
;    APPLICANT:  Kretschmer, Axel
;    TITLE OF INVENTION:  CDNAS CODING FOR MEMBERS OF THE
;    TITLE OF INVENTION:  CARCINOEMBRYONIC ANTIGEN FAMILY
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sprong Horn Kramer & Woods
;      STREET:  660 White Plains Road
;      CITY:  Tarrytown
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10591-5144
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 2.0 Mb storage
;      COMPUTER:  APPLE MACINTOSH 6500
;      OPERATING SYSTEM:  SYSTEM 7.5
;      SOFTWARE:  WordPerfect 6.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/468,856B
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/027,974
;      FILING DATE:  08-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/760,031
;      FILING DATE:  13-SEP-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/274,107
;      FILING DATE:  21-NOV-1988
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/207,678
;      FILING DATE:  16-JUN-1988
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/060,031
;      FILING DATE:  19-JUN-1987
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/016,684
;      FILING DATE:  19-FEB-1987
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 06/896,361
;      FILING DATE:  13-AUG-1986
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kurt G. Briscoe
;      REGISTRATION NUMBER:  34,141
;      REFERENCE/DOCKET NUMBER:  MEI 242.10 KGB
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (914) 332-1700
;      TELEFAX:  (914) 332-1844
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1591 nucleotides
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-468-856B-9


  Query Match                 0.7%;  Score 34.2;  DB 3;  Length 1591;
  Best Local Similarity   64.6%;  Pred. No. 4.1;
```

```
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/027,974
;       FILING DATE:  08-MAR-1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/760,031
;       FILING DATE:  13-SEP-1991
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/274,107
;       FILING DATE:  21-NOV-1988
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/[illegible]
;       FILING DATE:  16-JUN-1988
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/060,031
;       FILING DATE:  19-JUN-1987
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/016,683
;       FILING DATE:  19-FEB-1987
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 06/896,361
;       FILING DATE:  13-AUG-1986
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Kurt G. Briscoe
;       REGISTRATION NUMBER:  34,141
;       REFERENCE/DOCKET NUMBER:  MEI 242.9 KGB
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (914) 332-1700
;       TELEFAX:  (914) 332-1844
;   INFORMATION FOR SEQ ID NO:  8:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2009 nucleotides
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US-08-468-859A-8


  Query Match                 0.7%;  Score 34.4;  DB 4;  Length 2009;
  Best Local Similarity   54.8%;  Pred. No. 4.2;
  Matches   68;  Conservative    0;  Mismatches   56;  Indels    0;  Gaps    0;

Qy    4021 aatgcaatgtttcattgcttcaatgatcctgttgttgaaaatttccttgaa 4080
           ||||| | |       |  || |||| ||||||| || | ||||||| | ||  ||
Db     370 AGTGGAAGAGAAAGAGTATATTCCAATGCATCCCTGCTGATCCAGAATGTCACGCAGGAC 429

Qy    4081 aatgaaggaaactacgtct 4099
           ||| |||| ||||| ||
Db     388 gatgcagggatcctacact 406


RESULT  13
5169835-5
;Patent No. 5169835
;     APPLICANT: WAI-YEE, CHAN
;     TITLE OF INVENTION: PREGNANCY SPECIFIC PROTEINS AND APPLICATIONS
;     NUMBER OF SEQUENCES: 48
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/07/[illegible]
;       FILING DATE: 07-AUG-1989
;SEQ ID NO:5:
;       LENGTH: 1573
5169835-5


  Query Match                 0.7%;  Score 34.2;  DB 5;  Length 1573;
  Best Local Similarity   64.6%;  Pred. No. 4.1;
  Matches   51;  Conservative    0;  Mismatches   28;  Indels    0;  Gaps    0;

Qy    4 zl aatgcaaatgtttcattgcttcaatgatcctgttgttgaaaatttccttgaa 4080
```

Matches 51; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

Qy 4021 [illegible]

Db 436 AGTGGACGAGAAACAGTATATCCAAATGATCCCTCCGAATCCAGAATCTCACAACAAGAG 495

Qy 4081 [illegible]

Db 496 GATCCAGCATCCTACAACT 514

RESULT 15
US-08-408-859A-6
; Sequence 6, Application US/08408859A
; Patent No. 6022958
; GENERAL INFORMATION:
; APPLICANT: Barnett, Thomas; Ellner, James; Kmarick, Michael;
; APPLICANT: Kretschmer, Axel;
; TITLE OF INVENTION: DNAS CODING FOR MEMBERS OF THE
; TITLE OF INVENTION: PAPC IN EMBRYONIC ANTIGEN FAMILY
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sprung Horn Kramer & Woods
; STREET: 660 White Plains Road
; CITY: Tarrytown
; STATE: New York
; COUNTRY: USA
; ZIP: 10591-5144
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5in inch, 2.0 Mb storage
; COMPUTER: APPLE MACINTOSH 650
; OPERATING SYSTEM: SYSTEM 7.5
; SOFTWARE: WordPerfect 6.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/408,859A
; FILING DATE: 06 JUN 1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/027,974
; FILING DATE: 08 MAR 1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/760,031
; FILING DATE: 13 SEP 1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/741,107
; FILING DATE: 21 NOV 1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/537,678
; FILING DATE: 15 JUN 1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/460,031
; FILING DATE: 29 JUN 1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/310,688
; FILING DATE: 15 FEB 1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 06/899,361
; FILING DATE: 13 AUG 1986
; ATTORNEY/AGENT INFORMATION:
; NAME: Kurt G. Briscoe
; REGISTRATION NUMBER: 34,141
; REFERENCE/DOCKET NUMBER: MCI 24.2-6 KGB
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (914) 332-1700
; TELEFAX: (914) 332-1844
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1501 nucleotides
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-408-859A-6

Query Match 6.7%; Score 84.2; DB 4; Length 1501;
Best Local Similarity 64.6%; Pred. No. 4.1;
Matches 51; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

Qy 4021 [illegible]

Db 436 AGTGGACGAGAAACAGTATATCCAAATGATCCCTCCGAATCCAGAATCTCACAACAAGAG 495

Qy 4081 [illegible]

Db 496 GATCCAGCATCCTACAACT 514

Search completed: August 19, 2002, 09:44:09
Job time: 1121 sec

GenCore version 4.5
Copyright (c) 1993 - 2001 Compugen Ltd.

SW protein - protein search, using sw model

Run on: August 19, 2002, 13:44:11 ; Search time 67.00 Seconds
(without alignments)
4326.596 Million cell updates/sec

Title:
Perfect score: 7646
Sequence: 1 MASWTSPWWVLFGMVFMHSP.....LNLASLLRYKQKSVSSQE 1023

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172004929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_19:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

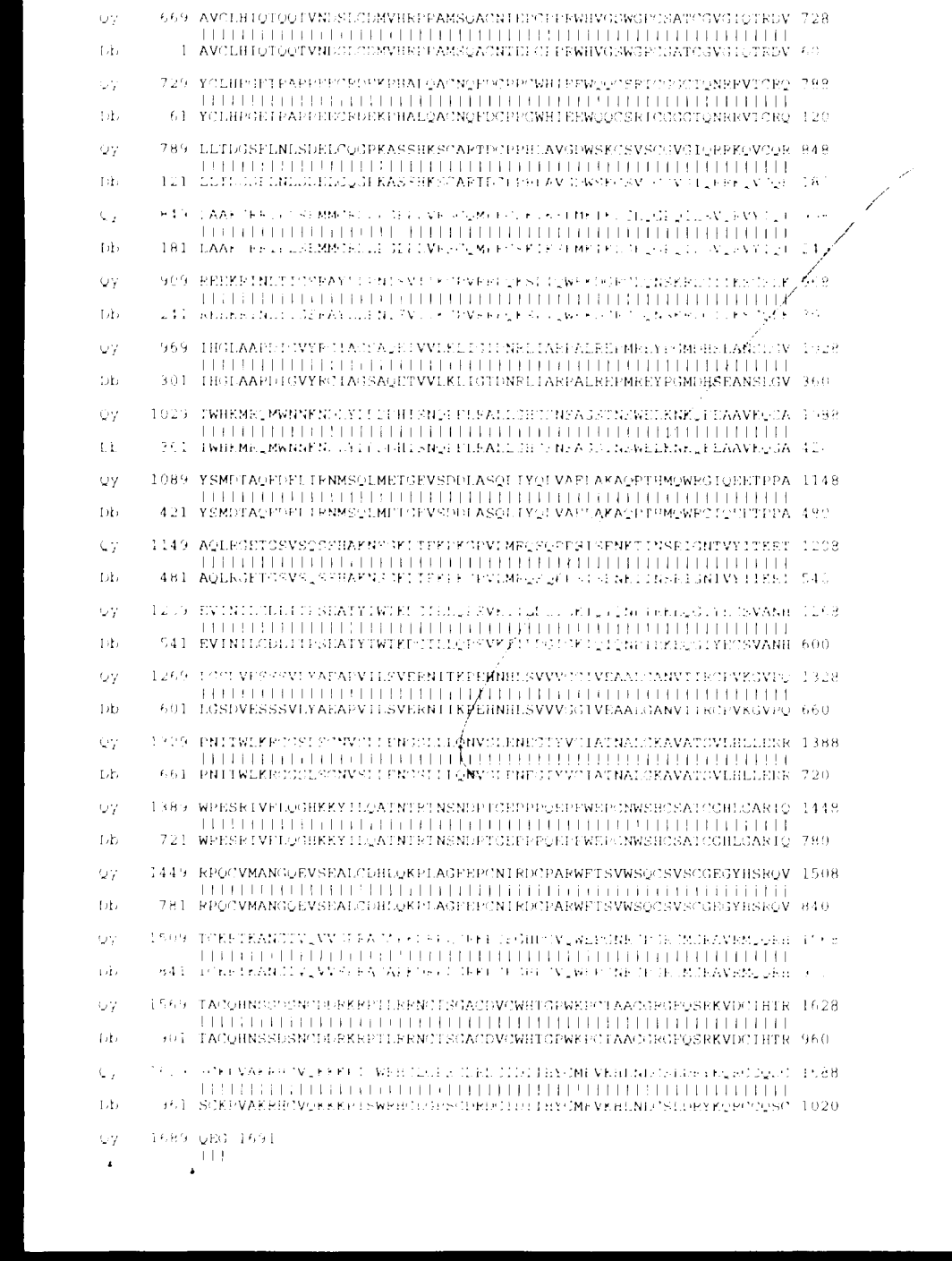
Result No.	Score	Query Match	Length	DB	ID	Description
1	5569	72.9	1023	4	Q90LI7	Q90LI7 homo sapien
2	4432	58.2	766	4	P82987	P82987 homo sapien
3	3729	48.8	761	6	Q95LQ2	Q95LQ2 macaca fasc
4	2018	26.4	525	4	Q96RW4	Q96RW4 homo sapien
5	1184	15.5	417	4	Q96BX8	Q96BX8 homo sapien
6	1170.5	15.3	1014	5	Q95J33	Q95J33 drosophila
7	1104.5	14.4	2155	5	Q17791	Q17791 caenorhabdi
8	1110.5	14.5	1054	5	P90884	P90884 caenorhabdi
9	958.5	12.5	1280	11	Q9EPX2	Q9EPX2 mus musculu
10	914.5	12.0	1455	4	Q96J28	Q96J28 homo sapien
11	898.5	11.8	2174	5	Q86JR0	Q86JR0 drosophila
12	891	11.7	3063	5	Q9VAV4	Q9VAV4 drosophila
13	869	11.4	2162	5	Q76864	Q76864 caenorhabdi
14	867.5	11.3	4109	5	Q86D6H	Q86D6H mansonia so
15	852.5	11.1	1572	5	Q44668	Q44668 tachaeleus
16	774	10.1	1427	4	Q96L37	Q96L37 homo sapien
17	771.5	10.1	858	5	Q9V9Q7	Q9V9Q7 drosophila
18	720	9.4	732	13	Q9EX59	Q9EX59 mus musculu
19	700	9.2	961	4	Q6U645	Q6U645 homo sapien
20	693	9.1	858	4	Q9UEZ4	Q9UEZ4 homo sapien
21	671.5	8.8	560	4	Q9NSJ8	Q9NSJ8 homo sapien
22	658	8.6	1054	5	Q9W493	Q9W493 drosophila
23	648.5	8.5	622	4	Q9HBX9	Q9HBX9 homo sapien
24	582.5	7.6	1229	5	Q9VE61	Q9VE61 drosophila
25	469	6.1	568	4	Q9BHE4	Q9BHE4 homo sapien
26	459	6.0	1444	5	Q17591	Q17591 caenorhabdi
27	438.5	5.7	1209	4	Q9UPZ6	Q9UPZ6 homo sapien
28	438.5	5.7	2674	4	Q96SC8	Q96SC8 homo sapien
29	438.5	5.7	3666	4	Q96RW7	Q96RW7 homo sapien
30	436	5.7	1536	4	Q96CH4	Q96CH4 homo sapien
31	434.5	5.6	454	4	Q96BB1	Q96BB1 homo sapien
32	434.5	5.6	1047	5	Q9XSV8	Q9XSV8 bos taurus
33	399.5	5.2	566	5	Q9VKV3	Q9VKV3 drosophila
34	378	4.9	4124	4	Q75B51	Q75B51 homo sapien
35	374.5	4.9	340	11	Q91Z56	Q91Z56 mus musculu
36	361.5	4.7	432	4	Q9NPM2	Q9NPM2 homo sapien
37	352	4.6	1062	5	Q19204	Q19204 caenorhabdi
38	345	4.5	1148	5	Q966Q0	Q966Q0 caenorhabdi
39	341.5	4.5	788	5	Q22641	Q22641 caenorhabdi
40	341	4.5	1121	5	Q966P9	Q966P9 caenorhabdi
41	320.5	4.2	244	11	Q9CBC7	Q9CBC7 mus musculu
42	311.5	4.1	685	4	Q9TTS5	Q9TTS5 bos taurus
43	292.5	3.8	446	1	Q43384	Q43384 homo sapien
44	291.5	3.8	364	4	Q9UQQ1	Q9UQQ1 homo sapien
45	291.5	3.8	769	5	O00816	O00816 toxoplasma

ALIGNMENTS

RESULT 1
Q90LI7
ID Q90LI7 PRELIMINARY; PRT; 1023 AA.
AC Q90LI7;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE KIAA1233 PROTEIN (FRAGMENT).
GN KIAA1233.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=20339619; PubMed=10774462;
RA Nagase T., Ishikawa K., Kikuno R., Hirosawa M., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XV.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 6:337-345(1999).
DR EMBL; AB033059; BAA86547.1; -.
DR InterPro; IPR003598; Ig_C2.
DR InterPro; IPR003006; Ig_like.
DR InterPro; IPR000006; Ig_MHC.
DR InterPro; IPR000884; ISPL.
DR Pfam; PF00047; ig; 3.
DR Pfam; PF00969; tspl; 1.
DR SMART; SM00408; IGc2; 2.
DR SMART; SM00419; IG_like; 1.
DR SMART; SM00209; ISPL; 1.
DR PROSITE; PS50835; ISPL; 1.
KW Immunoglobulin domain.
FT NON_TER 1 1
SQ SEQUENCE 1023 AA; 111521 MW; 10ABCFEFGAGAGIGAI CRC64;

Query Match 72.9%; Score 5569; DB 4; Length 1023;

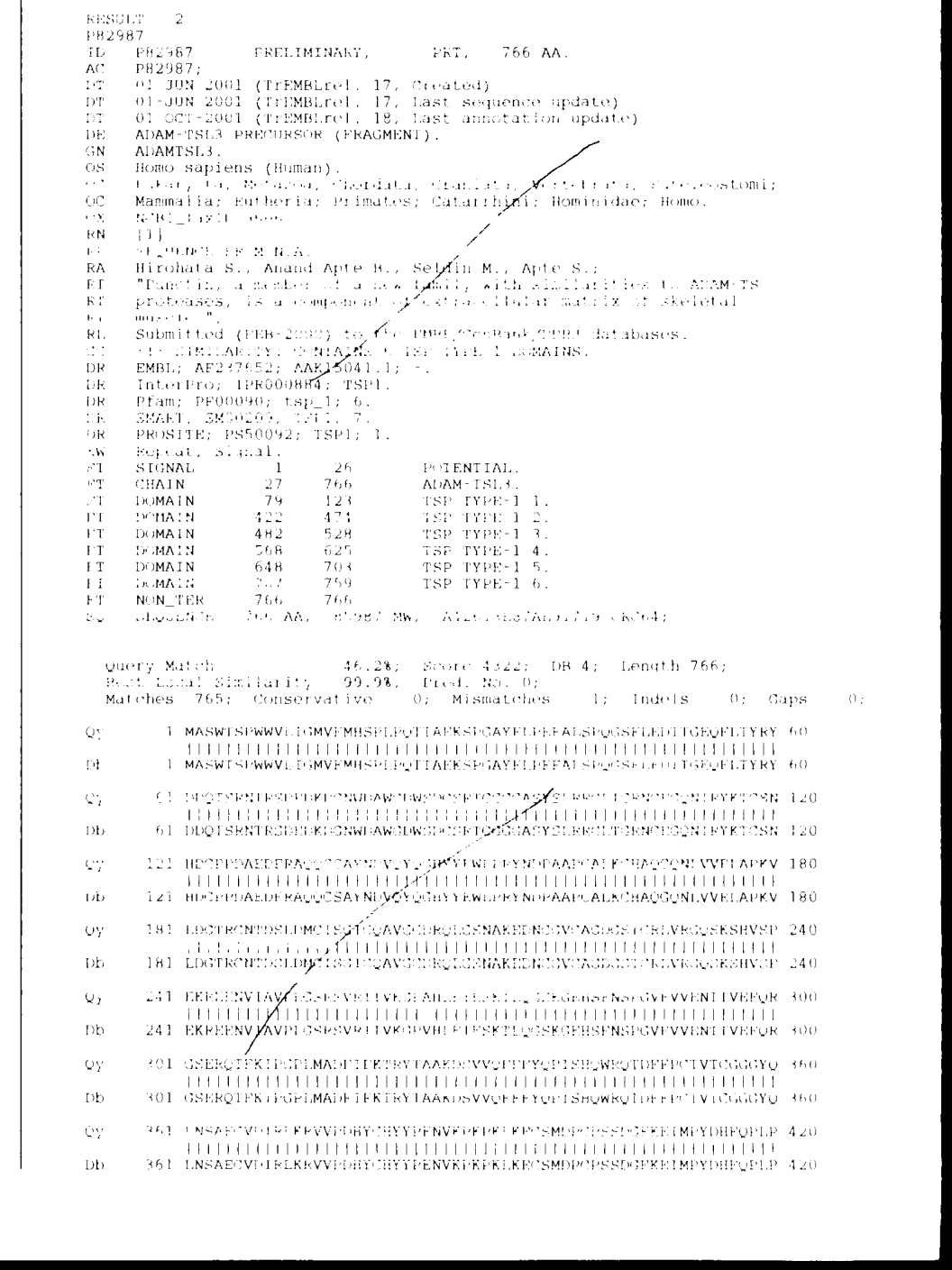
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1022; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

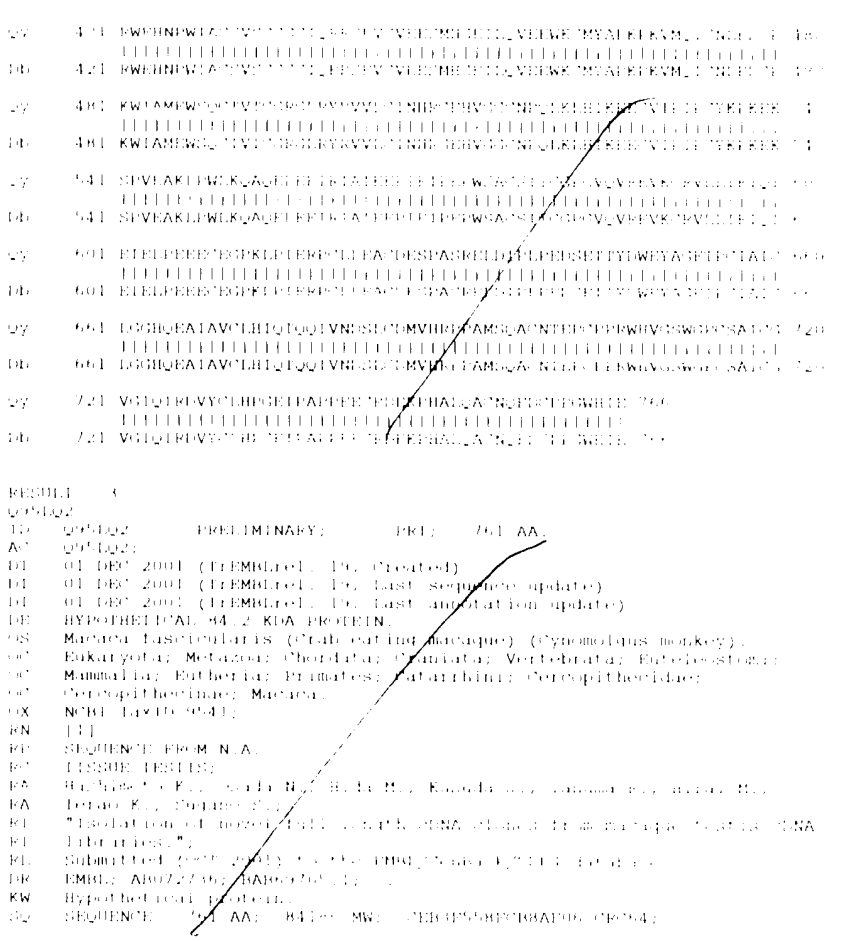


Db 1021 QEG 1023

RESULT 2
PB2987
ID PB2987 PRELIMINARY; PRT; 766 AA.
AC PB2987;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE ADAM-TSL3 PRECURSOR (FRAGMENT).
GN ADAMTSL3.
OS Homo sapiens (Human).
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RA Hirohata S., Anand Apte B., Sellin M., Apte S.;
RT "ADAM-TS, a member of a new family, with similarities to ADAM-TS
RT proteases, is a component of extracellular matrix of skeletal
RT muscle.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
RL Not in Medline; UI: [illegible].
DR EMBL; AF247652; AAK15041.1; -.
DR InterPro; IPR000884; TSP1.
DR Pfam; PF00090; tsp_1; 6.
DR SMART; SM00209; TSP1; 7.
DR PROSITE; PS50092; TSP1; 1.
KW Repeat, Signal.
FT SIGNAL 1 26 POTENTIAL.
FT CHAIN 27 766 ADAM-TSL3.
FT DOMAIN 79 123 TSP TYPE-1 1.
FT DOMAIN 422 474 TSP TYPE-1 2.
FT DOMAIN 482 528 TSP TYPE-1 3.
FT DOMAIN 568 625 TSP TYPE-1 4.
FT DOMAIN 648 703 TSP TYPE-1 5.
FT DOMAIN 707 759 TSP TYPE-1 6.
FT NON_TER 766 766
SQ [illegible] 766 AA; 87987 MW; A[illegible]/A[illegible] CRC64;

Query Match 46.2%; Score 4322; DB 4; Length 766;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 765; Conservative 0; Mismatches 1; Indels 0; Gaps 0;





RESULT 3
Q95QQ2
ID Q95QQ2 PRELIMINARY; PRT; 761 AA.
AC Q95QQ2;
DT 01-DEC-2001 (TrEMBLrel. 16, Created)
DT 01-DEC-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 16, Last annotation update)
DE HYPOTHETICAL 84.2 KDA PROTEIN.
OS Macaca fascicularis (Crab-eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC Cercopithecinae; Macaca.
OX NCBI_TaxID=9541;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RA Ha S.-Y., Park H.-J., Lee B.-H., Kim K., Kim J.-I., Lee J.-H.,
RA Jeong K., Park H.-S.;
RT "Isolation of novel full length cDNA clones from cynomolgus monkey testis cDNA
RT libraries.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AB072756; BAB69768.1; -.
KW Hypothetical protein.
SQ SEQUENCE 761 AA; 84144 MW; 7EB18F50B8C08A06 CRC64;

Query Match 21.9%; Score 2729; DB 4; Length 761;
Best Local Similarity 64.9%; Pred. No. 9.9e-283;
Matches 684; Conservative 17; Mismatches 207; Indels 0; Gaps 0;

Qy 940 QKSILQWLKLQMKLQNSKPLQTLKSQSLKLHCLAAIFLVVFLTASSAQLIVVLEKLLTL 999
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 27 EKSILQWLKLQMKLQNSKPLQTLKSQSLKLHCLAAIFLVVFLTASSAQLIVVLEKLLTL 86

Qy 1000 NKLIAPLALPFPMPLYSCMGLSLANSLCVIWHKMSQMWNNKNCLAYLCLHLSNQPFLSAL 1059
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 87 NKLIAPLALPFPMPLYSCMGLSLANSLCVIWHKMSQMWNNKNCLAYLCLHLSNQPFLSAL 146

Qy 1060 LGHCSNSAGSLNSWFLKNFQLFPAAVKQYAYGMGLALPLLLLGNMSLLMPLLVYFLLAYL 1119
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 147 LGHCSSSAGNLNSWFLKNFQLFPAAVKQYAYGMGLALPLLLLGNMSLLMPLLVYFLLAYL 206

Qy 1120 LYYLLVAPLAKAQFLHMQWWVLQEFMDVPAALIRGELCQVSQSNAKNSKELFKKKLVL 1179
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 207 LYYLLVAPLAKAQFLHMQWWVLQEFMDVPAALIRGELCQVSQSNAKNSKELFKKKLVL 266

Qy 1180 MKQSSLPRGSLENKLLNSKLNVYLKKLEVNELSLELLELGAYYWKKNHSLLPEVK 1236
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 267 MKQSSLPRGSLENKLLNSKLNVYLKKLEVNELSLELLELGAYYWKKNHSLLPEVK 323

Qy 1237 LLLSLDKLQTLNELEELSLYSLQVANHLSLGAVLSSGVLTALAIVLLSVLENLLKFEHN 1296
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 324 LLLSLDKLQTLNELEELSLYSLQVANHLSLGAVLSSGVLTALAIVLLSVLENLLKFEHN 383

Qy 1297 HLSVVVSSLVLAALSANVTEFLVKSVFQFENTWLKKVLKLGNVSLLFNFSSLLLQNVSL 1356
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 384 HLSVMVSSLVLAALSANVTEFLVKSVFQFENTWLKKVLKLGNVSLLFNFSSLLLQNVSL 443

Qy 1357 ENSCLYYCIAINALSKAVAIFVLELLPFWPLSKIVFLQEKKYILQAINLFLNCNDFLG 1416
||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 444 ENSGLYYCIAINALSKAVAIFVLELLPFWPLSKIVFLQEKKYILQAINLFLNCNDFLG 503

Qy 1417 FEFLLSFDFWLSLSWSCSATCQELGARIQKFQCPVMANQQEVSEALCIELLKPLVGPQPCN 1476
|| ||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 504 FEFLPQFEFPWEPSNWSRCSATCQELGARIQKFQCPVMANQQEVSEALCIELLKPLVGPQPCN 563

Qy 1477 IKQFSARWFLSVWSQLSVSCGSGLHSRQVPCRRFKANGIVQVVSPPACAPKPRPCLRRPC 1536
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 564 IKQFSARWFLSMWSQLSVSCGSGLHSRQVPCRQFKANGIVQVVSPPACAPKPRPCSRRPC 623

Qy 1537 LGLPVGLLGGLGGLSPKVLPLLRNCISGACDV 1599
|||||||||||||||| |||||| ||||||| |||||||||| ||||||||||
Db 624 FGMPVQWLPQNQFSGRMGMAVRMQQHLTACQHNGSSGNCDSRKRFLERNCISGACDV 686

Qy 1600 CWHIGPWKPCTAACGRGVQSRKVLCIRTRSCRPVAKRP VQKKKPCSWKPCLGPSCDRDC 1659
||| |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db 687 CWHAGPWKPCTAACGRGLQSRKVLCIRTRSCRPVAKRP VQKKKPCSWKPCLGESCDSIY 746

Qy 1660 I 1660
|
Db 747 I 747

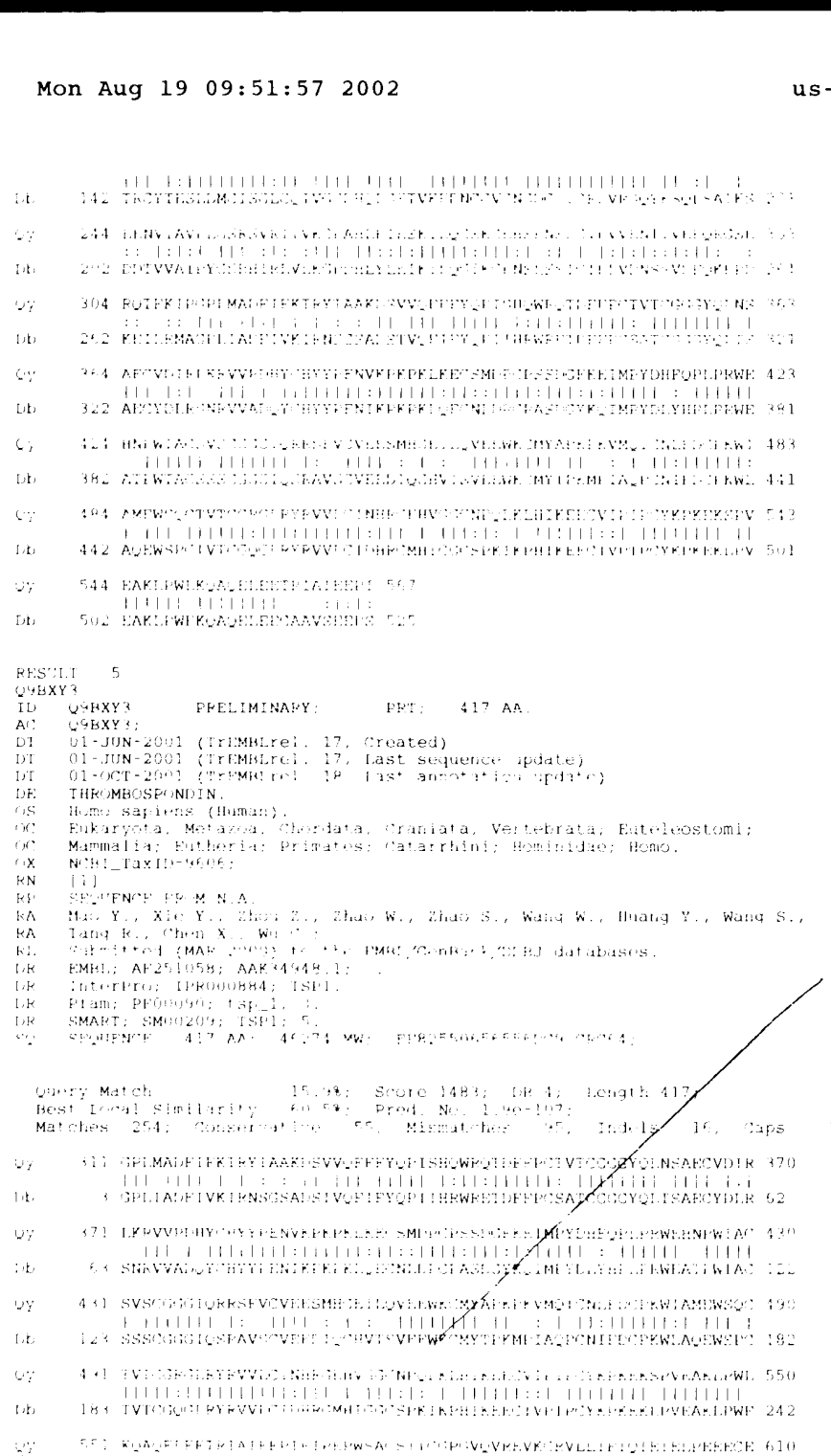
RESULT 4
Q96RW4
ID Q96RW4 PRELIMINARY; PRT; 525 AA.
AC Q96RW4;
DT 01-DEC-2001 (TrEMBLrel. 16, Created)
DT 01-DEC-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 16, Last annotation update)
DE ADAM 15 RELATED PROTEIN.
GN ADAM15RP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Hirohata S., Abe C.S.;
RT "A novel member of ADAM 15 related gene, ADAM 15RP (A disintegrin like
RT And Metalloproteinase domain with ThromboSpondin type 1 modules
RT Related gene 1).";
RL Submitted (AUG-1999) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF176515; AAK84170.1; -.
SQ SEQUENCE 525 AA; 58351 MW; 6F180C7CAAA7BA0B CRC64;

Query Match 21.9%; Score 2048; DB 4; Length 525;
Best Local Similarity 68.9%; Pred. No. 1.2e-209;
Matches 544; Conservative 72; Mismatches 89; Indels 0; Gaps 0;

Qy 94 LGSNGRGGLGKGNGWGAWGWGWGGGGGAGYGLGNKQLGRNGTGLGNGYKFGSNHD 153
||| |||||||| |||| ||||||||||| |||||| |||||||||| |||||||||| ||
Db 22 LGKIARGRLGRGLGWGAWGWGWSGGGGGAANGLGNKQLGRNGTGLGNGYKFGSNVD 81

Qy 154 YDAEDFRAQGKQTGAYNGAVGYGTPDFESWGLKYNDLAAFGALKCHAGGNGVVELAPKVLDG 213
|||| |||||||||| |||||||||| |||||||||| |||||||||| |||||||||| ||
Db 82 YDAEDFRAQGKQTGAHNGAVKHGGPEYEWGLKYSNDFSNFGGLKCHAGGNGVVELAPKVLDG 141

Qy 214 IPNTLGLEMKTGVQTVAVTEGAGNAKELNKTVTAGQNLPIVEGSGKCHVSEEKK 267

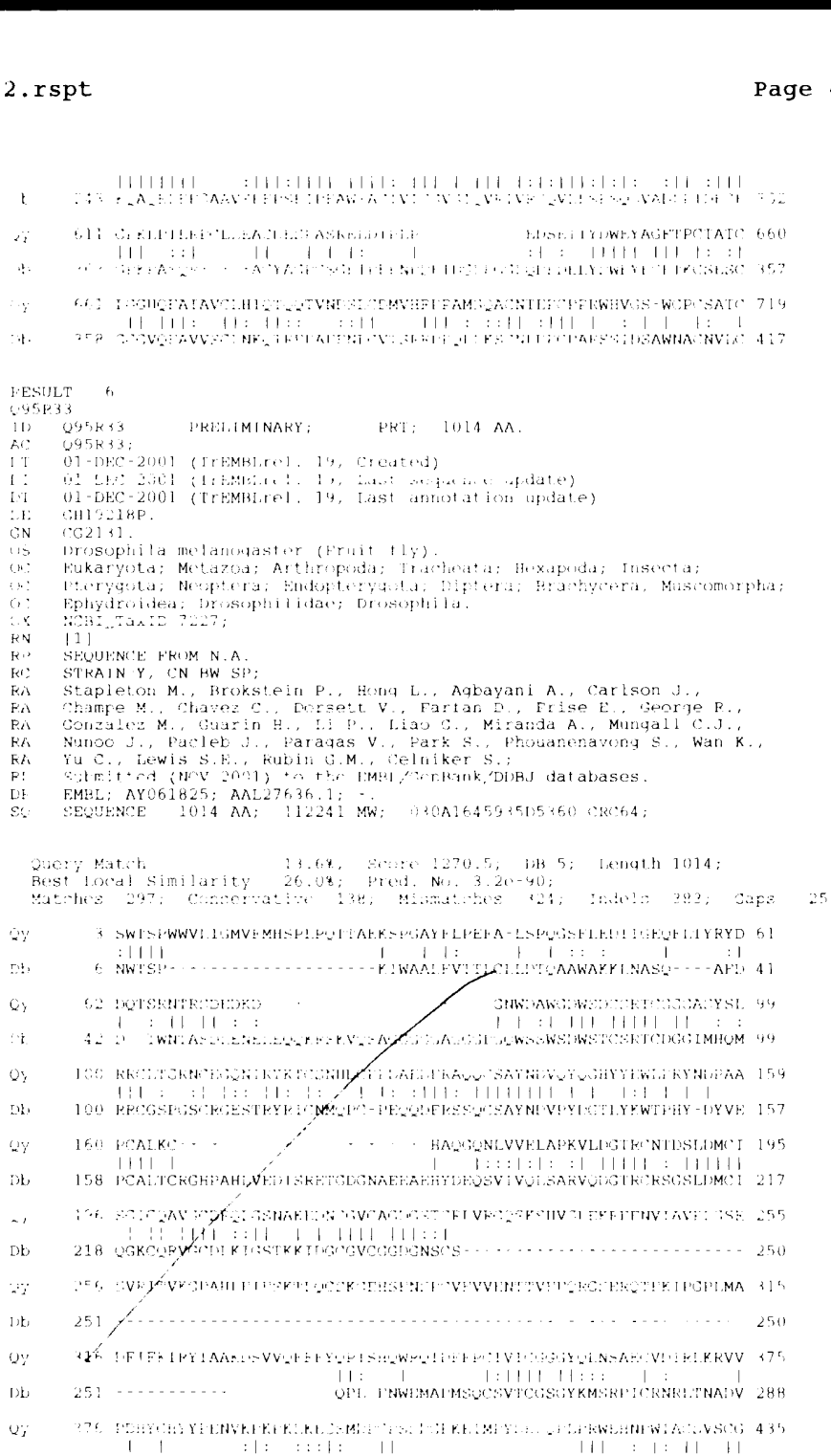


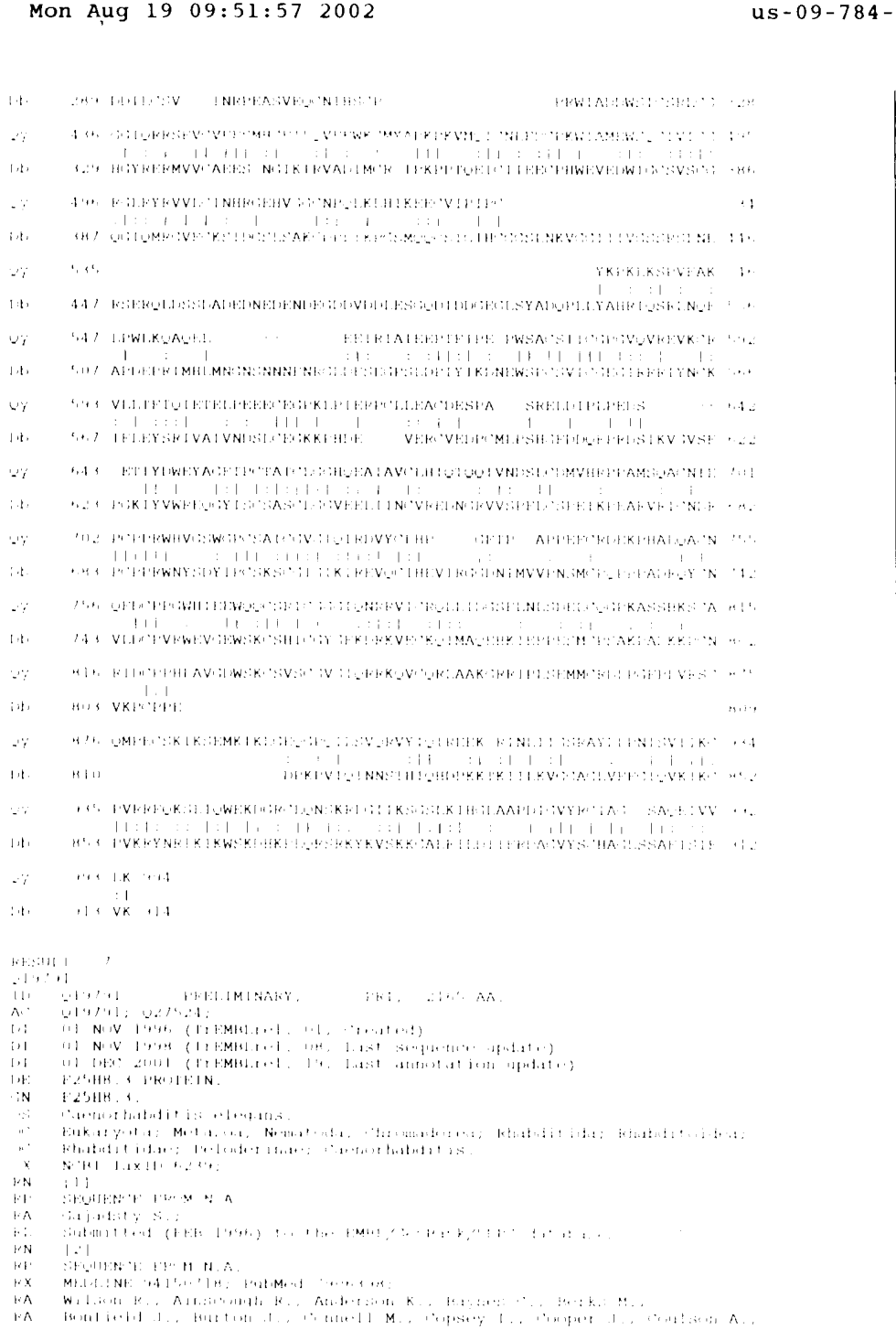
RESULT 5
Q9BXY3
ID Q9BXY3 PRELIMINARY; PRT; 417 AA.
AC Q9BXY3;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE THROMBOSPONDIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Mao Y., Xie Y., Zhou Z., Zhao W., Zhao S., Wang W., Huang Y., Wang S.,
RA Tang R., Chen X., Wu C.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF251058; AAK34648.1; -.
DR InterPro; IPR000884; TSP1.
DR Pfam; PF00090; tsp_1; 3.
DR SMART; SM00209; TSP1; 5.
SQ SEQUENCE 417 AA; 46274 MW; FF8F5506F5F6F4C9 CRC64;

Query Match 15.9%; Score 1483; DB 4; Length 417;
Best Local Similarity 60.5%; Pred. No. 1.9e-107;
Matches 294; Conservative 55; Mismatches 95; Indels 16; Gaps 3;

RESULT 6
Q95R33
ID Q95R33 PRELIMINARY; PRT; 1014 AA.
AC Q95R33;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01 DEC 2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE GH19218P.
GN CG2141.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Y, CN_HW_SP;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Fartan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao C., Miranda A., Mungall C.J.,
RA Nunoo J., Pacleb J., Paragas V., Park S., Phouanenavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL Submitted (NOV 2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AY061825; AAL27636.1; -.
SQ SEQUENCE 1014 AA; 112241 MW; 030A16459345D5460 CRC64;

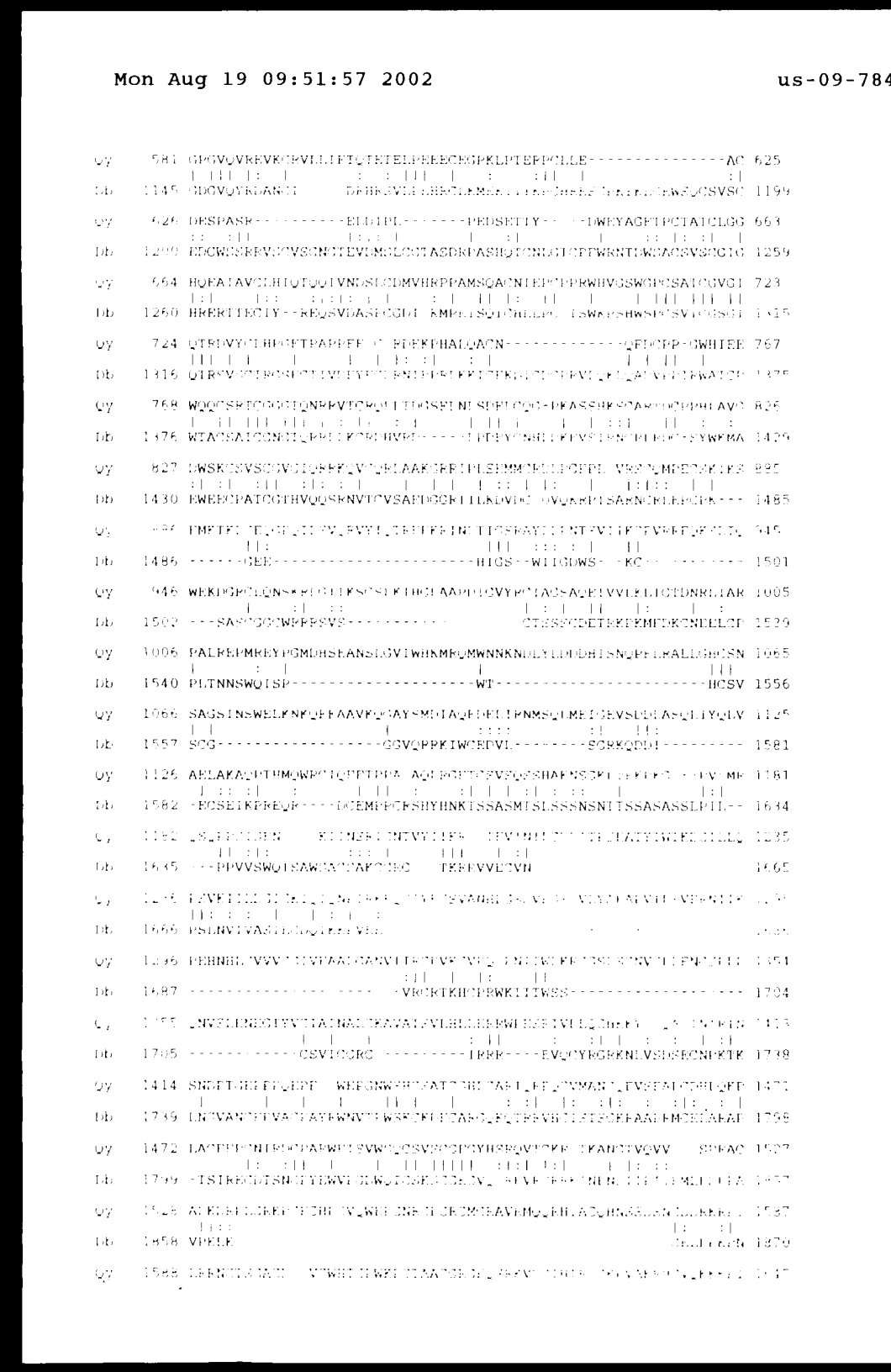
Query Match 13.6%; Score 1270.5; DB 5; Length 1014;
Best Local Similarity 26.0%; Pred. No. 3.2e-90;
Matches 297; Conservative 138; Mismatches 421; Indels 282; Gaps 25;





RESULT 7
Z19791
ID Z19791 PRELIMINARY; PRT; 2165 AA.
AC Z19791; Q27524;
DT 01-NOV-1996 (Rel. 44, Created)
DT 01-NOV-1998 (Rel. 56, Last sequence update)
DT 01-DEC-2001 (Rel. 76, Last annotation update)
DE F25H8.3 PROTEIN.
GN F25H8.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94150716; PubMed=7906600;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laisster N., Latreille P.,
RA Lightfoot J., Lloyd C., Mcmurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken J., Roopra A., Saunders D., Shownkeen R.,
RA Sims M., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
DR EMBL; Z69361; CAA93288.1; -.
DR EMBL; Z69360; CAA93288.1; JOINED.
DR EMBL; Z69360; CAA93287.1; -.
DR EMBL; Z69361; CAA93287.1; JOINED.
DR HSSP; P15167; 1DPH.
DR MEROPS; M12.145; -.
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000884; TSP1.
DR InterPro; IPR000140; Zn MIpeptidse.
DR Pfam; PF01421; Reprolysin; 1.
DR Pfam; PF00090; tsp-1; 14.
DR SMART; SM00209; TSP1; 18.
DR PROSITE; PS50215; ADAM_MEPRO; 1.
DR PROSITE; PS50092; TSP1; 6.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN 1.
SQ SEQUENCE 2165 AA; 244897 MW; F0C4DA8AAA9C4888 CRC64;

Query Match 12.7%; Score 1188.5; DB 5; Length 2165;
Best Local Similarity 22.7%; Pred. No. 4.6e-85;
Matches 405; Conservative 213; Mismatches 578; Indels 58; Gaps 8;



Qy 581 GPGVQVKEVKGRVLLIFTGTETELPEEECMGPKLPTEPPCLLE----------AC 625
Db 1145 GDGVQYKDANGI DFHNSVLLGHCLMEA----CHEEGWAWLCEWGQCSVSC 1199

Qy 626 DESPASR----------ELLIPL-------PEDSETTY-- --DWEYAGFTPCTAICLGG 663
Db 1200 EDGWSSREVGCVSCNCTEVDMSLCGTASDRPASHQTCNLGTCFTWRNTDWCAGSVGGTG 1259

Qy 664 HQFATAVCLHTQTQQIVNDSLCDMVHRPPAMSQACNIEPTTPRWHVGSWGPCSATCGVGT 723
Db 1260 HRERTTECIY--REQSVDASFCGDI KMPLISQTCNLLPC TSWKASHWSPCSVTCGSGT 1315

Qy 724 QTRDVYCIHFGFTPADPFE C FDEKPHALQACN----------QFDCPP-GWHIEE 767
Db 1316 QIRSVYCTRQSDFTIVEYFC ENETPKLFPTCPKL CTCTPEVLGKLCASYTFDWAICP 1375

Qy 768 WQQCSRTCGGGIQNRPVICRQLTTGSPLNLSPELCQG-PKASSRKSCAKHCPPHLAVG 825
Db 1376 WTACSAICGNGIQKRDIKCRQHVPD-----IPEEYCNGLKFEVSPCNLPLCPYWFMA 1429

Qy 827 DWSKCSVSCGVGIQSRPVVCRQLAAKGRPIPLSIMMCKGLPCPPL VRGCNMPEDCPKIKR 885
Db 1430 EWEECPATCGTHVQQSKNVTCVSAPDGGRILDDVDC QVQARPISARNCFLEPCPA--- 1485

Qy 886 EMPTPC TEVGPVLEVFLTPEFPINTIGSRAYCLNTPVCIFTEVWRFLPKCLL 940
Db 1486 ------GEE---------------------HIGS--WTIGDWS- -KC-- -------- 1501

Qy 941 WEKIGPCLQNSRRICLKSCSLKTHCLAAPDIGVYPCIACSAQEIVVLELIGTDNRLIAR 1000
Db 1502 ---SASCGGCWPPPSVS------------ CTSSPCDETRRPEMFDKCNEELCP 1539

Qy 1001 PALREPMKEYPGMDHSEANSIGVIWHKMRQMWNNKNNDLYLPDDHISNQFFLKALIGHCSN 1060
Db 1540 PLTNNSWQISP--------------------WT----------------------HCSV 1556

Qy 1061 SAGSINSWELKNFQFFAAVFVCAYSMDIAQFDELIPNMSQLMEICEVSDDLASQLIYQLV 1120
Db 1557 SCG-----------------GGVQPPKIWCEDVL---------SGRRQDDI---------- 1581

Qy 1121 AELAKAQPTHMQWPCTQPTTPPA AQIPCGTCSFVSQQSHAFNSSFLSFFTPC --FVCMF 1181
Db 1582 -ECSEIKPREQK----DCEMPPCFSHYHNKISSASMISLSSSNSNIISSASASSLPIL-- 1634

Qy 1182 SSSFFCLSEN EEIMEAICNETVYIIPS CFVTMDIC---TEIAIYIGICFCIGLL 1235
Db 1635 ---PPVVSWQISAWCATCAPCIGC TKRFVVECVN 1665

Qy 1236 EVPFIIL IISSSSSWSWSWSFFSSVANEISSVSS VSVSIAFVIF-VFSNIPF 1289
Db 1666 PSLNVIVASIISSSSSFFFVFL 1686

Qy 1290 PEHNELCVVVCIVFAACCACVIFFCIVFCVCL INICMCFFCTSCSCVCIIPNFCIII 1351
Db 1687 ---------- -VRCRTKHCPRWKLIIWSS---------- 1704

Qy 1352 SNVSIENESIFVCTIAINALIFAAVAIFVLHLIIFRWIFSSIVIILCDLFS SSVFRIS 1413
Db 1705 ----------CSVICGRC ---------IRRR----EVQCYRGRKNLVSDSECNPKTK 1748

Qy 1414 SNGFTGIFFCIDPF WEFGNWFRCFATCNICIACFIFCVMANICFVSFAICTCHICPF 1473
Db 1749 LNCVANCTFVACIAYFWNVCIWSRCNLICAACCLFCTRFVBCIISICRFAAIFMCSICAAAF 1808

Qy 1474 LACTPFCNIFSCCPARWFISVWSCCSVSCSICYHSFQVFCKF CKANCTVQVV SSFFAC 1533
Db 1809 -ISIRRCDVTSNCIYIWVICSWCTCSFCCCDCNC SIVFCRRCNLRCIICMCLCIIA 1867

Qy 1534 ACKCRFCLCREFTCHF CVWFICNFCRCMCRAVFMCCFHLACCHNSSCCCLCRRRC 1587
Db 1868 VPFLF 1879

Qy 1588 LRRNCLCIACC VCWFICWFCCAACCCCCFRCVC CCCCC CCCCCCCCCCCCC 1647

Db 1871 ESQTCSLNPCSPSFDFWSFICFWSCSCRNTCQCCIPSFRVFCVCNNCPPVFPVFCT TFFPR 1928

Qy 1648 CWRHCLCPSCDRDCTDFTHYCMFVKHLNLCSLD 1678
Db 1929 RTQYCF----EKNCLPST--CQFIKSQNVFAKD 1955

RESULT 8
P90884
ID P90884 PRELIMINARY; PRT; 1059 AA.
AC P90884;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-JAN-1999 (TrEMBLrel. 09, Last sequence update)
DT 01 DEC 2001 (TrEMBLrel. 19, Last annotation update)
DE F53B6.2 PROTEIN.
GN F53B6.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA White S.;
RL Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; Z81086; CAB03121.1; -.
DR InterPro; IPR003598; Ig_c2.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR000884; TSP1.
DR Pfam; PF00047; ig; 1.
DR SMART; SM00408; IGc2; 1.
DR SMART; SM00209; TSP1; 8.
DR PROSITE; PS50092; TSP1; 2.
KW Immunoglobulin domain.
SQ SEQUENCE 1059 AA; 117768 MW; 5B182A42E41C8597 CRC64;

Query Match 21.8%; Score 1113.5; DB 5; Length 1059;
Best Local Similarity 21.1%; Pred. No. 1.1e-77;
Matches 351; Conservative 170; Mismatches 451; Indels 695; Gaps 52;

Qy 77 NWLAWSWLLSLLLGSAVAGFLFLMSLVAMWRVDSKRHFTPGAELEEAQQCS 136
Db 25 SWAGWSWLWSLFSLSSAAGSLAFSLSSLSLGVRLAVTAGPLSLSAGLAGSLGS 84

Qy 137 AVNCVQVGCPVYFWLTRVNLFAAFCAIFSPAL QCLEVVFLLAFEVLFLTESCN TDSLDM 196
Db 84 G EEIVSRCQ CEVVCRSRLTGANFLW RVDIRTPCQAATSRAV 124

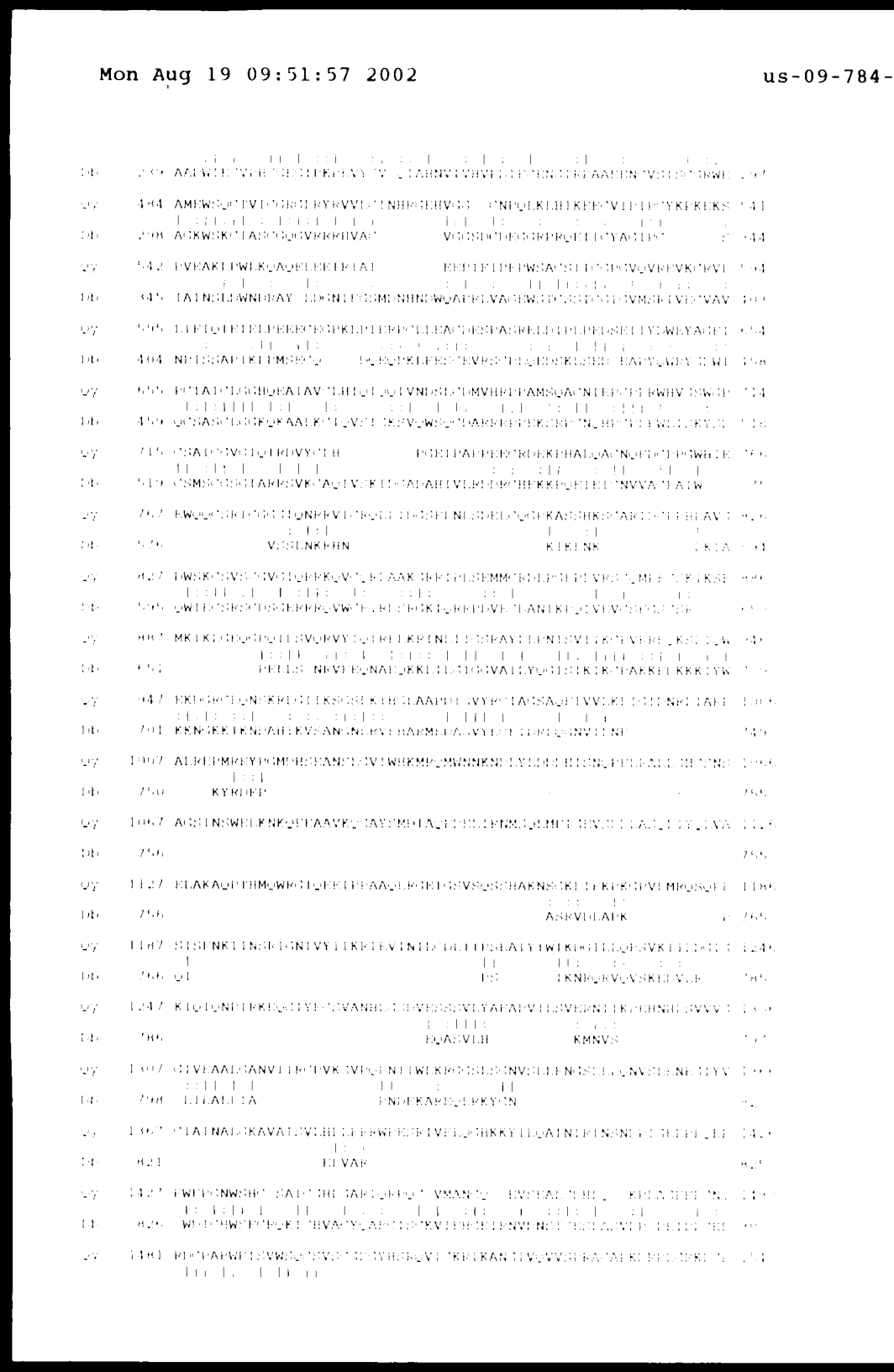
Qy 197 SLSLSLSLAVLSLSLSLSMARLFNCIVSLSL SELRVSLSERSVSLERRLRAVIAVSLS 256
Db 125 CSKGSCQIVSCDCDLISSSFRRFDACGVCGGRGDTC---------- 158

Qy 257 SESVFITVFGFAMLFILSFLFQSGRGEDSINSFSVFVVFLNFIVLFQSSFLFQTFKIPGPL 316
Db 159 ---------- ---------- ---------- ---------- ----- 160

Qy 317 MAPSIPKT RYTAAK- ESVVLPTFTLISHQWFCTTTPTCIVTCGCTYQLNSAE--- 365
Db 161 -SEIWKVLLATACACN LKIV LW SSAGRSIASTSQPI 198

Qy 366 ---VLLKEVVLLLLSSSLLSAVKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK 423
Db 199 VVCVNA GEVVPLKLC ADELKEVLAKLDMLTC VLLWKCMYAGKPKVMQLNLST IEWI 258

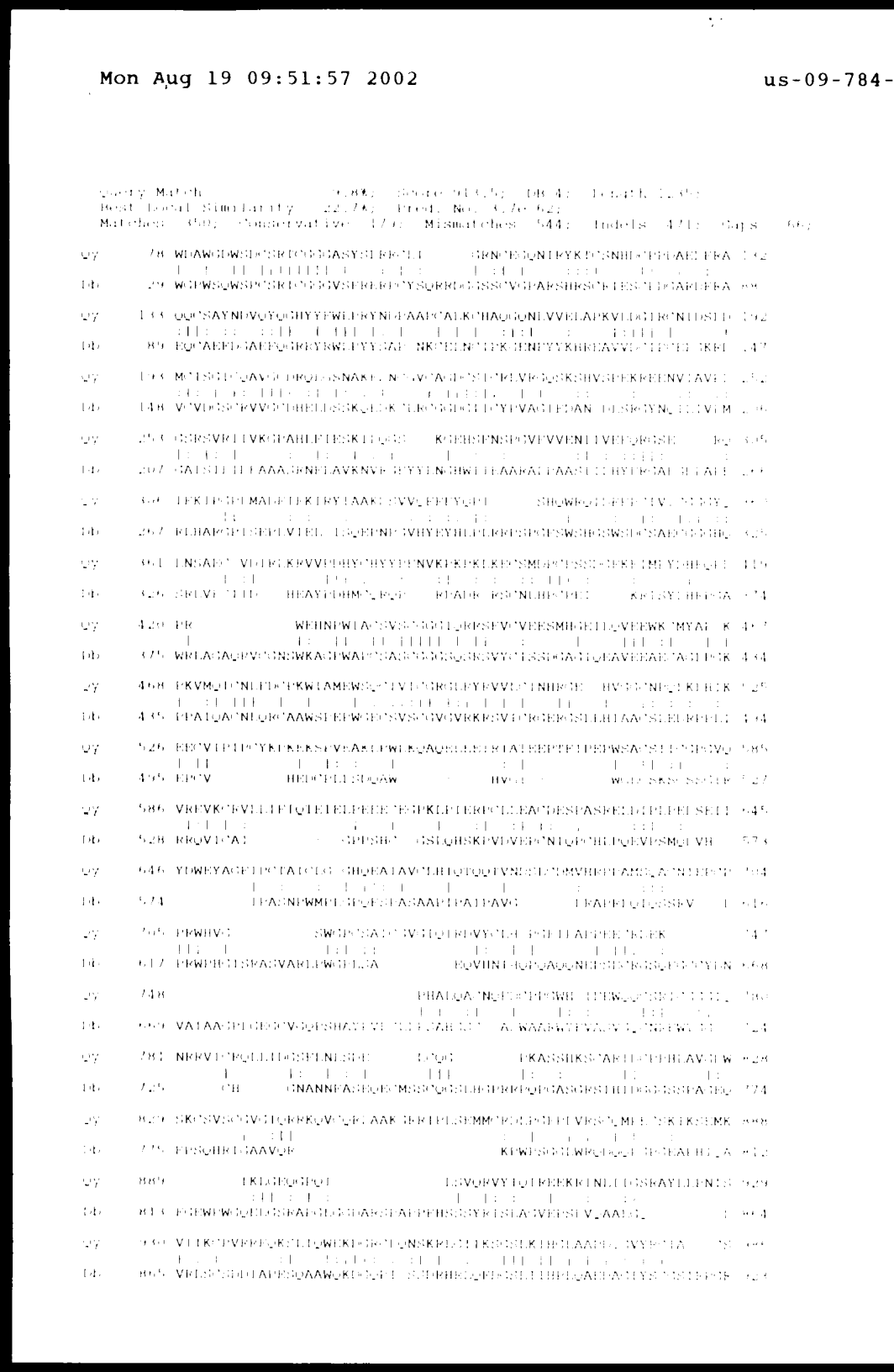
Qy 424 SNFWIACSVSSIIILLKLLVSVSLSMDLSLVLWKCMYAGKPKVMQLNLSTIEWI 483



RESULT 9
Q9EPX2
ID Q9EPX2 PRELIMINARY; PRT; 1280 AA.
AC Q9EPX2;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE PAPILLIN.
GN MUS musculus (Mouse).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL;
RX MEDLINE=20534499; PubMed=11062047;
RA Kruse M.-N., Becker C., Lauchart W., Recknagel D., Plewig G.,
RA Kruczek-Kaminska S., Bornkamm G.W., Wolf K., Sterry W.,
RA Kramer M.D., Lommatzsch J., Kaufmann R., Huber M., Hamm H.,
RA Brocker E.-B., Zollner T.M.;
RT "Expression of papillin, the human orthologue of chicken
RT basement membrane protein 200, in epithelial tissues and
RT inflammatory cells.";
RL J. Invest. Dermatol. 115:971-976(2000).
DR EMBL; AF314171; AAG41980.1; -.
DR HSSP; P12111; 2KNI.
DR InterPro; IPR003599; Ig.
DR InterPro; IPR003598; Ig_2.
DR InterPro; IPR003600; Ig_like.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR002223; Kunitz_BPTI.
DR InterPro; IPR000884; TSP1.
DR Pfam; PF00047; ig; 3.
DR Pfam; PF00014; Kunitz_BPTI; 1.
DR Pfam; PF00090; tsp_1; 5.
DR PRINTS; PR00759; BASICPTASE.
DR SMART; SM00409; IG; 3.
DR SMART; SM00408; IGc2; 4.
DR SMART; SM00410; IG_like; 2.
DR SMART; SM00131; KU; 1.
DR SMART; SM00209; TSP1; 5.
DR PROSITE; PS00280; BPTI_KUNITZ_1; 1.
DR PROSITE; PS50279; BPTI_KUNITZ_2; 1.
DR PROSITE; PS50092; TSP1; 5.
KW Immunoglobulin domain; Serine protease inhibitor.
SQ SEQUENCE 1280 AA; 140824 MW; AE2877C5E56IAF40 CRC64;

Query Match 22.1%; Score 958.5; DB 11; Length 1280;
Best Local Similarity 22.1%; Pred. No. 1.2e-05;
Matches 478; Conservative 178; Mismatches 690; Indels 545; Gaps 62;

RESULT 10
O95428
ID O95428 PRELIMINARY; PRT; 1235 AA.
AC O95428;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 133.5 KDA PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Rowen L., Madan A., Qin S., Abbasi N., Dors M., Ratcliffe A.,
RA Madan A., Dickhoff R., Shaffer T., James K., Lasky S., Hood L.;
RT "Complete sequence of the gene for presenilin 1.";
RL Submitted (NOV-1998) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF109907; AAC97963.1; -.
DR HSSP; P12111; 2KNT.
DR InterPro; IPR003598; Ig_c2.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR002223; Kunitz_BPTI.
DR InterPro; IPR000884; TSP1.
DR Pfam; PF07647; ig; 3.
DR Pfam; PF00014; Kunitz_BPTI; 1.
DR Pfam; PF00090; tsp_1; 5.
DR PRINTS; PR00759; BASICPTASE.
DR SMART; SM00408; IGc2; 4.
DR SMART; SM00131; KU; 1.
DR SMART; SM00209; TSP1; 5.
DR PROSITE; PS00280; BPTI_KUNITZ_1; 1.
DR PROSITE; PS50092; TSP1; 4.
KW Hypothetical protein; Immunoglobulin domain;
KW Serine protease inhibitor.
SQ SEQUENCE 1235 AA; 142477 MW; A6B44A0E4F38E350 CRC64;



```
DR   PRINTS; PR00759; BASICPHASE.
DR   SMART; SM00409; IG; 3.
DR   SMART; SM00408; IGc2; 2.
DR   SMART; SM00410; IG_like; 1.
DR   SMART; SM00131; KU; 3.
DR   SMART; SM00280; I5PI; 7.
DR   SMART; SM00217; WAP; 1.
DR   PROSITE; PS50279; BPTI_KUNITZ_2; 3.
DR   PROSITE; PS50092; I5PI; 5.
KW   Matrix protein; Serine protease inhibitor; Signal.
FT   SIGNAL        1     26       POTENTIAL.
SQ   SEQUENCE   2174 AA;  [illegible] MW;  038F7079525A3120 CRC64;


  Query Match             [illegible]%;  Score 898.5;  DB 5;  Length 2174;
  Best Local Similarity   24.8%;  Pred. No. 1.4e-60;
  Matches  314;  Conservative  164;  Mismatches  467;  Indels  473;  Gaps   48;

Qy      10 VLIGMVFMHSFLFQVITAEKSIGA-YFFLPEFALSPQQSFLEDITGEQFLIYRYDDQTISRNI  68
           ||  :    |  |   :  ||  :|||  :::| |
Db      17 VLACIHDSQCHFPSLRQKFQVYCANHYLIEGSVIPGGE-  -----      --  -  --- ---  53

Qy      69 EGDEDLKDGNWSAWGDWSQCSRTCGGGASYSLRECLTGRN---  -TSGQNIRYKTCNNHDC  123
            :| |:    |  |   |||||||||| ||  | ||    :      | | : |: :|:  ||
Db      54 GNDPDE---WTPWSSPSDCSRTCGGGVSYQTRECLRRDDRGEAVCSGGSRKYFSCNTQDC  110

Qy     124 PPDAEDFRAQQCSAYNDVQYQGHYYEWLPRYNDPAAPCALKCHAQGQNLVVELAPKVLDG  183
           | :  |||||||| ::   |: | :||||:  | |  || | |  :|:           |||||
Db     111 PEEESDFRAQQCSRFDRQPTGCVFYTWAF  NECELNGCMPKGDKFYYRQREKVVDG  169

Qy     184 TRCNTFGSIMKTSGTCQAVCRNKQGIRSNAKPVNCGVDGSTCPKIVPQQ-SESGNVSPEK  242
           |||| |||:|:| |  |||| |||:|||| |  | ||||||: :|   :  ::::|
Db     170 TRCNDKDLDVCVNCSCMPVCCDMMLGSDAKLDRCPKCGGDGSTCKTTRNTITTRNLAPGY  229

Qy     243 PPPNVIAVPIGSSPSVPRTTVPSTAHIPTISPPTTQCS---PCPPHSPNSTCVTVVPNTIVEP  299
            :  :: :| |: ::||    |:  :  :   |     |:   : |       |:   :|
Db     230 NG  LLLLPSCAINIPTRETVESSNYLACPNRSCGHYTLRCLWRTLPTRTMPTARSWWNTQ  287

Qy     300 PGSPRQTPKIPGPIMAIPIPKTRYTAAKDSVVQPPPVQPISHQWPQTIPPPCTVTCCCCY  359
           |                    || |                              ||| |
Db     288 R-----------------  --KPMGPAAPPQ-----------------  ------LTCSGPI  405

Qy     360 QLNIAIPVTVTGLPWVSLMSLPVPLNVPLTPLKLLSMPCPSLLPGPLIMPLCNPLVSMGLL  419
           :    : :: | :  | | :  |||:                                   |
Db     406 SESIFTIVMIVQPPNISIP  VPVPTIPPSISHPQ-                        457  [illegible]

Qy     420 PPWPHNPWTACSVSCGCCILPLPFVCVPPSMHGPTLQVPPWPCHVAIPKPKVMQIPCNLPDC  479
           | |: : ||| ||||| | |   |  :    : :|    |  ||  | |    |
Db     340 HTWTHHQPNACSASCGGGSQNKKVIC---NNRITLAPVNPSLCPQKSKPVPEQACCIPPC  396

Qy     480  PPWIAMPWPLTVIIGPF SLAYSTVCA  INRRCHRVIICNPLSLPLHPEEPCVCIICPP  5 5
           | |:  ||||:|  ||  | : : :  |  |:  |||          :|:|  |    |
Db     397 APHWVPLGEWSPCSKTCGCGIMQPLNPSITCPRPSGGPEH          VPELAV    GL  442

Qy     546 PPPPPSPVSAPLPWPLQAQPLPPIPASPPTPTPPPWSASSIPCPPVQVPPPVPCPRVLL  605
           |     |      |  : :  |     | :   ||: |   |: | | | :|  |    :
Db     443 PPVIGPT         ---APPLIGPPVPPVTPPAPPWPVPPPPPPPPPPPPPVT  :: ::

Qy     606 TFTQTPTPLPPPPCPGPKLPTPPPCLLPACPPSPASPPLDIPLPPPSPTTYPW-  -PYAC  662
           :  ||||:|   |  ||: ||| |:                          ||   :::|
Db     492 PPNGHPPVLTPPPLPVPPKPPTPPPSCLLIPCP-----------------CVIWIISVWPC  542

Qy     663 FTPC ----IAICLGGHQEAIAVCLHIQIQQTVNDSICDMVHRPFAMSQACNIPPCPPKW  717
           |     | | : |:|:             :   :  |:     | :|: | :  |  :|
Db     543 CNAPCLPNTIPIAICGPPPPL          PVTPLPPCP  PPVPPLPPPPPPPPPPPPW  593

Qy     718 HVGSWCPCSATCAVGLPKRAVTCIRSGP       TPAPPP-PCPPPP-PPPAPLPACNPPP-  775
               | ||| || |:::| | :   ||      |||  : :|  | || : | |   :
Db     594 PPPLWPPPPAPPPPVPPPVTPPPPPPP  -PPLPPPIVTPPALPPPPPPNPPPPPPPPPPW  652

Qy     776 -PPPWHIPWVLVSPTCTPPTNPPVIPPLLTPGPTLPLPPPPPLPCPKASPHPCPART  832
            ||  :   | :|:| |:|:  :  | |              | | |
Db     653 VPPPWHIPWVLVSPTCTPPTNPPVIPPLLTPGPTLPLPPPPPLPCPKASPHPCPART  709

Qy     818 DCPPHLAVCDWSKCSVSCGVGIQRRKQVVQRLAAKGRRIPLSEM    MCRDLPGPPLVR  873
                 |||:|                  : :  ||||     | :   ||
Db     672 --------------SVNCD  --  --          EEKVIEPLSEKCNGEACTEDEIILPLTS  703

Qy     874 SCQMPFTPFIFSPMFIFLTPQPQIISVQFVYQIQVKFKPNIIYSRAVTFNFTVTVK  928
           :        |   : :    | | :::|   :    :  | |:|   ::    |   :::
Db     704 T      EFECIGPELEPCDLQCLLLIC  CDLSCLEPELVLDLCLAPIPLEPAEDLMQ  756

Qy     934 ----GVRRPQKSIQWFKNGKPLANSKRIIFKPNSLRIHSAAPSDVGVFQCWAQE  989
               |   :  :|  |   :::   ||: ||                  |  :|
Db     757 SDSPTPYDEFTFTGTTPEIGSGYDSESTTDSGISIEGS             GDDEE  798

Qy     990 IVVLELICIFLNELIARLALELMRLKY MCRRFANPLDPWWNPNLNDLYLDDDH  1049
           |                  |||::                ||    |
Db     799 T                  SEASI                DLSSSTDS  842

Qy    1050 ISNQPPLEALICPETNCAVTCTNFWELENFQFTAAVPQCAYGNNTAQIPELLRNHSQLMET  1109
           |              |: : |: | :  ::   ::|   : | | :             ||
Db     843 GSISS------DSISSDSSSSISSDAISEAPASSVSDSSDSIDASP----------  ET  894

Qy    1110 GFVSPPLASQLIYQIVAFLAFAIQPTHMQWPQTQFPT--------PPAAQLPGETGSVSQS  1161
           ||||          | : :: | |  |  | :|          | :    : :  :
Db     855 TGVSDDSTD---VSSSTPASASPSTDVS  GASDSTGSINASDSIPFSSTFASSSTIDSI  909

Qy    1162 SHAPNCQPIIPPPPQVLPPQQLPPPPPSPPLPPAPPINAVLPPVPVNIPPVPPPS  1221
           : ||          |    ::  | | : : :|  |:|   ::   ||           |
Db     910 DSSDNSSN--------VSEGSTEASSSSVSDSSDSSDGSTDGVSSTTE-------  NSS  952

Qy    1222 EAIYIWIPIPIIIPIPVPIIIIPIPIIPIIPVIIPPPPPPPPPPPPPPPPPPPPPPPPPPPPP  1277
             ::|  | | |    |  | :   :: |  |    : :|    |  ||||
Db     953 DSTSDATSDSTASSDSTDSTSDQTIETITPESSTDSTESSTILDASSTIDASSTSESSS  1009


RESULT  12
Q9VAV4
ID   Q9VAV4      [illegible]      PRT;  3060 AA.
AC   Q9VAV4; Q9VAV3;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   CG11540 PROTEIN.
GN   [illegible]
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   [illegible]
RX   MEDLINE 20196006; PubMed 10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   [illegible],
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
```

```
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Lin X., Mattei M., McIntosh I.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Sidén-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.-A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Zmasek H.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS, A AND B (SIMILAR TO BOTH);
CC       SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
DR   EMBL; AE003765; AAF56765.2; -.
DR   EMBL; AE003765; AAF56795.2; -.
DR   HSSP; P12111; 2KNI.
DR   FlyBase; FBgn0003432; Sps.
DR   InterPro; IPR000561; EGF_like.
DR   InterPro; IPR000498; Ig_C2.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR002223; Kunitz_BPTI.
DR   InterPro; IPR000884; TSPI.
DR   InterPro; IPR002221; WAP.
DR   Pfam; PF00047; ig; 3.
DR   Pfam; PF00014; Kunitz_BPTI; 1.
DR   Pfam; PF00090; tsp_1; 5.
DR   Pfam; PF00095; wap; 1.
DR   PRINTS; PR00759; BASICPTASE.
DR   SMART; SM00408; IGc2; 3.
DR   SMART; SM00131; KU; 1.
DR   SMART; SM00209; TSPI; 7.
DR   SMART; SM00217; WAP; 1.
DR   PROSITE; PS00317; 4_DISULFIDE_CORE; 1.
DR   PROSITE; PS50280; BPTI_KUNITZ_1; 1.
DR   PROSITE; PS51229; BPTI_KUNITZ_2; 1.
DR   PROSITE; PS00022; EGF_1; UNKNOWN 1.
DR   PROSITE; PS50092; TSPI; 5.
KW   Alternative splicing; Immunoglobulin domain;
KW   Serine protease inhibitor.
FT   VARSPLIC    2003    2003       S -> SVVEV (IN SHORT ISOFORM).
FT   VARSPLIC    2044    2056       ENEKIMEDSCII -> VASPPLHFNAV (IN SHORT
FT                                  ISOFORM).
FT   VARSPLIC    2057    6064       MISSING (IN SHORT ISOFORM).
SQ   SEQUENCE   6060 AA;  687779 MW;  A7A51D3EE58C7259 CRC64;

  Query Match          96.5%;  Score 4011;  DB 5;  Length 6060;
  Best Local Similarity   95.8%;  Pred. No. 8.2e-260;
  Matches  807;  Conservative  10;  Mismatches  45;  Indels  37;  Gaps  47;

Qy      44 FLPEFALSPQGSPLEDCIGPQFTYRYIDQQIGRNIRSDEGKDCNWLAWSDWSPCSRCTGG 103
           ||||.|||| |                 :|| |   |  |   |.||||||||  |
Db       2 YLPESGVIPQGE                 GNDPDE   WTPWSSPSDCSRCTGG 35

Qy     104 GASYSLRPCTLIGRN      TEQCNTRYKPCCNHDCPDAEDFRAQQCSAYNGVCYGGHYY 148
           |.|| | ||  :    | | :  || :|  ||| :  |||||||| :  : | | :|
Db      36 GVSYQLRDCLKHDGCEAVCSGGRRYFSCNTQDCLEEECDPRAQDCSRPCDCQGTCVEY  9

Qy     149 EWLPRYNDPAAFCALKCRAQCNLVVEGLAEVCSCPNNSCGLMSLSCCQAVPTDPQCL 208
           |||| | |  || || |  :|          |||||||||  |||||||| | |||  :
Db      96 EWVPFYINAPNPCELNCMPEGSISISVLEEVVCSCPNNKCGLMSLSCCQAVPTDPQCL 154

Qy     209 GSNAKEENCTGVCACDSSCPLVRSQ GKSHVSPEKPEENVIAVPLSSKGSVRLLVKSPAHL 267
           |:.|||| | ||||||:  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db     155 GSGAKECKCRKCQGDGSCKEIPNCIIKGLAPCYNP  IELLEPCAINEKPEIVEGSN 212

Qy     268 FTDSKTDQGS  KDEHSFNSPSVIEVVENTIVEPQPSSEPQTEKTPSTEMASTEPKTEVS 327

Db     213 YLACRNHSCHIYYLNGDWRTDPPRPMFFANSWWNYQR--  --  --  --  --  -KPMGF 254

Qy     328 AAKDSVVQPFFYQCTSHQWRQTDPFPCTIVTCGGGYQLNSADCVDIEKKRVVFCHYCHYYP 387
           || |                        :|| |   :  :  :  :  :  | |  :  |
Db     255 AAIDQ                      LDCSGFISESLFIVMLVQEKNISLD YEYSIP 289

Qy     388 ENVEPKPPKLKLSMEGTSIFLEGDMPYIHPQFPLPEWEHNFWLAVSVSQQSLQPPSFV 444
           |::                                |    | |:  :  :|||| | |
Db     290 EGLGHSQ                          QDIHIWIHHQFNAVSASQQQSQSGKKVI 324

Qy     445 CVLDLMPEGGLVELAPKMSAGRKVPLCNGEP  NGEP  GGLAPSWLCVL SPCGLKYRV 502
           |    :  :|    |  ||  | |    | |  :  |||||  ||  | :  |
Db     325 NNKTILALVNFGLCQAQGRFVLLQAQSILPA AHWVLGLMGKKDKSQGGLAPQNKS 381

Qy     503 VLPCINHPGEHVGCNPQLKQHIKELCVCPPHCYKFKLKSFVLAKLPWLKQAQECEELR 560
           :|  |: |||          :|:|  |    | |          |        | :  |
Db     382 LPPLLSGGNAI          VLLGAV   CLKLVGNAI          ALKQECNR 417

Qy     561 IAIEEPTEIPEPWSACSIDCGPGVQVRLVRCRVLDLQPEIEETEEETGPKLPLERPC 620
           |  :  |||  | |||| | |:|  |  :  :  |||||  |  |  | :|
Db     418 DVKNCPSKAIGCPWIPCDKIPGRCQDPKKYCP  ELLNCHKRYVLPLPC VELKPETEKSC 475

Qy     621 LLEACDESPASKELLIPLPEDSELIYDW  EYAGFIPC      IAICLSGHQEAIAVCL 672
           |:  |:                    ||   :::|  |      | | : |:::|
Db     476 LLIPCE                  SVDWIISQWSGCNACGQNIEIRKIAICGNKPG      512

Qy     673 HIQTQQTVNDSLCDMVIKPEAMSQACNIEPCTPRWHVGSWGPCSATCGVGIQTRDVYCLH 732
           :  :  :|  | |||| | :|  |  :  :|  | | :|||| :|  :|  |
Db     513      EVIPPPPCE  PEVPSLSPPCESPCPEACWLSSPWSPCSATCGVGSKIVD      564

Qy     733 PQE      TPAPPE PCRDECKPHAIQACNQETC CPPGWHTEEWQQCSKSCGGGTQNR 782
           :        ||| : :|  | :| : | |  :  :| :  | :|| :|||| |||| : |
Db     565 CGLPQCPIVPEAPPVSMC NSECPPDSPQVPCSGPRVCPPPWFCGWSECSKPCGGGIRVR 622

Qy     783 KVICVQLLDGSELNLSDLLQQRKASSHKSCARICCPPHLAVCGWSKCSVSCGVCIQRR 842
           | |                | | :                          ||:|
Db     623 EVLC                ESNCIK                      SVNCD      637

Qy     843 KQVCQELAARSPKTLSEM     MCKSLPGPLVKSCQMPDCSKIKSEMKIKGGPCPQCI 898
           :  :  ||||    |  :   ||  :  |  :  :  :  :  :  :  :  :  :  :
Db     638           ELEAVELLGELSNCNLAC IECLLLLLESV       ENPIECCEECCGLIATEL 682

Qy     899 LSVQKVYLQIREEKPINLIIDSRAYLIFNETVLIK    CPVPEEQKSELQWEKIGKCLQ 954
           ||   :    :  |  |:|   ::    |    :::      |   |:  :   :|  |  :
Db     683 LSIGLSGEEKGEEVTILDSIAKTLLIIEAIEGLMQLSGSLIFYLPPLSICLPLGLSVLGP 741

Qy     955 NSKKDQIIKKDGSLKLHGLAAPLTGVYPCIAGSAQEIVVELKLIGIDNKLIAKPALKEPMKE 1014
           ::   ||:  ||                 |  :||                        |
Db     742 SIIDSGIGIEGS                 GDDEEI                        759

Qy     1015 YPGMSHCEANCTGVTWHKMPCMWNNKNCQLYLPCHICNCPETLALCPCTNCACCLNCWE 1074
            ||||:                  ||   |  |                  | : :  | :
Db     760        SEAAI                DLSSSIDSGSISS      IGISSDSSSGISSD 791

Qy     1075 LKNKQFEAAVEQAYSMGIAQGTELIPNMSQLMEIGEVSDDLASQLIYQLVAELAKAQPI 1134
              ::  :::|  : : | :                 ||  ||||           |  :: |
Db     792 AISEAPASSVSDSSDSIDAS      IETIGVSDDSID   VSGSIEAASESI 836

Qy     1135 HMQWRGIQEET          PPAAQLPQEIGSVSQSSHAKNSGIKLCPKPKGPVLMPQSQPP 1194
            :  | :|              | :  :  :  :  :  :  :  :  :  |  ::
Db     837 IVS   GASDSIGSINASDSIPESSIEASSSIDDSIDSSDNSSN         VSESSIEAS 886

Qy     1195 STSPNKLINSKTDNIVYTIKPTEVINILTCLLTISLAIYIWKRQCILLQPCVKTLLDGIG 1254
            | | : :|  |:|  :  :|              :::|  :  :  |  |
Db     887 SSSVSDSSNDSSDGSIDGVSSIIE          NSSDSIGDAIGESSAASESSIDSGSDQII 937

Qy     1255 KLQLQNPIPKKEQGIYEPCSVANIEGSDVESSS 1287
            :  :: |  |    : |       | ||||
Db     938 EIIEESSIGSIESSIILASSIGASSISESSS 969

RESULT  10
```

```
076840
ID   76840      PRELIMINARY;      PRT;   2167 AA.
AC   076840; Q22911;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 16, Last annotation update)
DE   C37C3.6 PROTEIN.
GN   C37C3.6.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadoria; Rhabditida; Rhabditina;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Celniker S., Bradshaw H.;
RT   "The sequence of C. elegans cosmid C37C3.";
RL   Submitted (FEB-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- ALTERNATIVE PRODUCTS: TWO FORMS (A AND B) MAY BE PRODUCED BY
CC       ALTERNATIVE SPLICING OF THE SAME GENE. THE SEQUENCE SHOWN IS THAT
CC       OF FORM B.
DR   EMBL; U64857; AAC25868.1; -.
DR   EMBL; U64857; AAC25867.1; -.
DR   HSSP; P00591; 1CHE.
DR   InterPro; IPR004598; Ig_c2.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR002223; Kunitz_BPTI.
DR   InterPro; IPR000884; TSP1.
DR   Pfam; PF00047; ig; 1.
DR   Pfam; PF00014; Kunitz_BPTI; 11.
DR   Pfam; PF00090; tsp_1; 6.
DR   PRINTS; PR00759; BASICPTASE.
DR   SMART; SM00408; IGc2; 1.
DR   SMART; SM00131; KU; 11.
DR   SMART; SM00209; TSP1; 7.
DR   PROSITE; PS00280; BPTI_KUNITZ_1; 10.
DR   PROSITE; PS50279; BPTI_KUNITZ_2; 11.
DR   PROSITE; PS50092; TSP1; 4.
KW   Alternative splicing; Hypothetical protein; Immunoglobulin domain.
KW   Serine protease inhibitor.
FT   VARSPLIC    147    155       MISSING (IN ISOFORM A).
FT   VARSPLIC   1556   1558       KDD -> SKF (IN ISOFORM A).
FT   VARSPLIC   1559   2167       MISSING (IN ISOFORM A).
SQ   SEQUENCE   2167 AA;  237599 MW;  9627478D5B32639 CRC64;


  Query Match             9.2%;  Score 863;  DB 5;  Length 2167;
  Best Local Similarity  28.6%;  Pred. No. 7.8e-58;
  Matches  220;  Conservative  114;  Mismatches  312;  Indels  122;  Gaps   30;

Qy      45 LPEFAISPQGSFLEDFIGEQFL----------------  -TYRYDDQTSRNTRSDED 73
           :| :|  |   | :|  | :|                    |   | :   :| :
Db      15 VPTWAFSLFFFFSFVFVFFFFFFFFFA  -VFF-FFFFVFVFVFVFVVFFFFFF-F 71

Qy      74 KDCNWLAWGDWSDCSRTCGGAGYQLRRCLDGRNFDGQNIEYKYCNHDKTPDAIGPAAC 133
           : |||  |   ::|||:|:|:|     ::|    :| | :|:|| :|: : |     ||||:
Db      72 EIGNWGHWVPLNDCSRRCGGGVQLTFRQC  SGPFPGASVRFLYCDLNAC DSTTFPAH 131

Qy     134 QCSATNQVLFLVDWSYSMNDRLAAVKQALKLFIRAL-QNCVVSLAAFNDSAQLVFPNT 192
           ||| :|:|   |:::::| | |      | | |  :  |   : | |||:|||:|:: | |:
Db     132 QCSKFNDPALMCNYHFWIP YEC KNPCELVCKPEC GNEYYEWALKVVGTRCVSKSNE 191

Qy     193 CPSPIVAVCVLPPLPSNAKPFNCQVCACPVTCPELVPQL SPCPVVSPPPPPPNVIAVP 252
           |: | |  |||| :|||: | | || | |||||||: : |:  : :::||   :  :| :|
Db     192 CVIGECLPVGCLVGEGGSSLEFFRTCCPPGSTCRLIECPPPEPNLSPCYHD--FIKLPE 247

Qy     253 CSPSVPLTVPVDPAHLEIFSPTLQCSPVSEHSTPNSPCVPEVVCNT -   TVEPLRCCEPCTP 307
           :: :::|     |     :  |:        |  |: ||       |:     :| :|
Db     248 GAINIKIQ   PARKSTNNLALKNCSDHFYLNGNELIQVIKEVEVCGTTIPVYDDAIPITL 307

Qy     308 KIPGPLMAIFIFFIFV-IAAFISVVLFFFYLR----ISHLWPLVKIPDIVDDQYLLN 362
           |||  :    :  :::| ::: |  |    : : ::  :: |||:|:|| | |
Db     308 SAQGPLSEELIVALIPPPSMCIAIVVFFSVIFIFFFVFVYMVFFFNKIPGVAVPPVVLAN 367

Qy     363 PAECVKQTIELKEVVFDHVYHHTYPDRVFEFEKL KEVCMDETFCTCSLIKEIMPYDHFQPLPR 421
           :  |:| : |  | |  |     ||   ||:  | |     | :
Db     368 NLVTIPGKNECDVLIKNFL  ENNAIKTFIFRESGEIVDAFEA----------------F 404

Qy     422 WFHNFWFACSVFCGGIQQRSVECVRSSNFLLIQVELWPCMVAIKKPKVMQTCNLFDCP 480
           |    | :|| :||  | | |    | :   :|  : ||:  |    :| | |||| | ||
Db     405 WFTGDWEPCSSTCGDGSQQQYRVVYCHQVFANGRKVIVEEHNFT VLKPEVKQTCNRFACP 464

Qy     481 EWGAWSWCPVPCDMGVQVRSRSVVCIRSNFNVCPNLVKEIRPCNVPCPVYEKELK 540
           :| | ||  |:  ||   :|| | |  |  | |    : ||    : |
Db     465 EWQAGPWSACSKTCGDAFQYRSVICMKSSKKG---EDGKLLAADAC-- ---------- 505

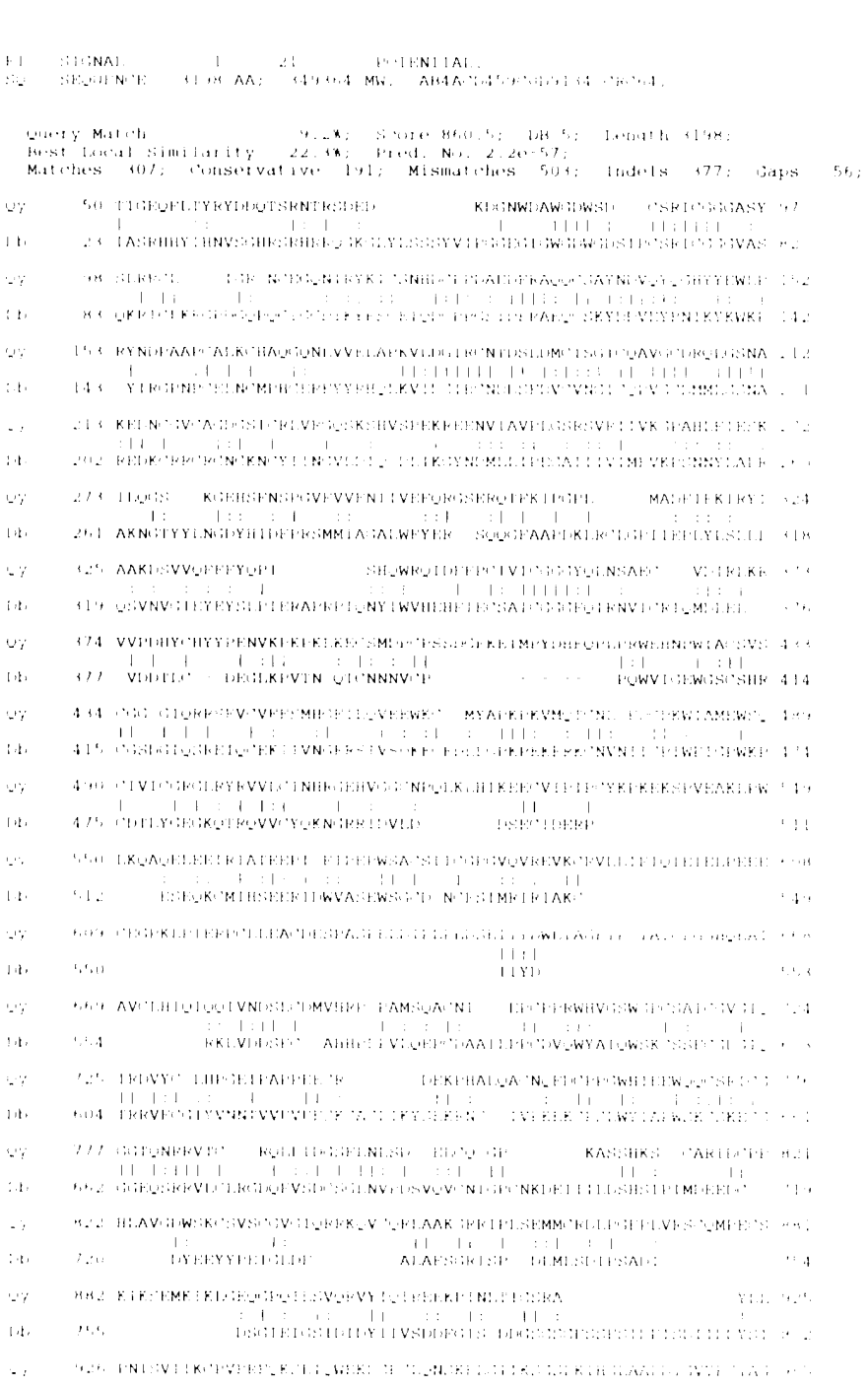
Qy     541 SPVPAKLPWLKQAQFLPPTPIATEPP -- TFTPPPWSAPSTFPGPGVQVPEEVKCRVLLT 596
           | :      : :: :  |     |    ||:   |: | | |     : ||| |:
Db     506 PAD       PLTKPTTHKTCNLCPTSCLTPVTCHWNLT IKCNDTEETFRPVTCK      552

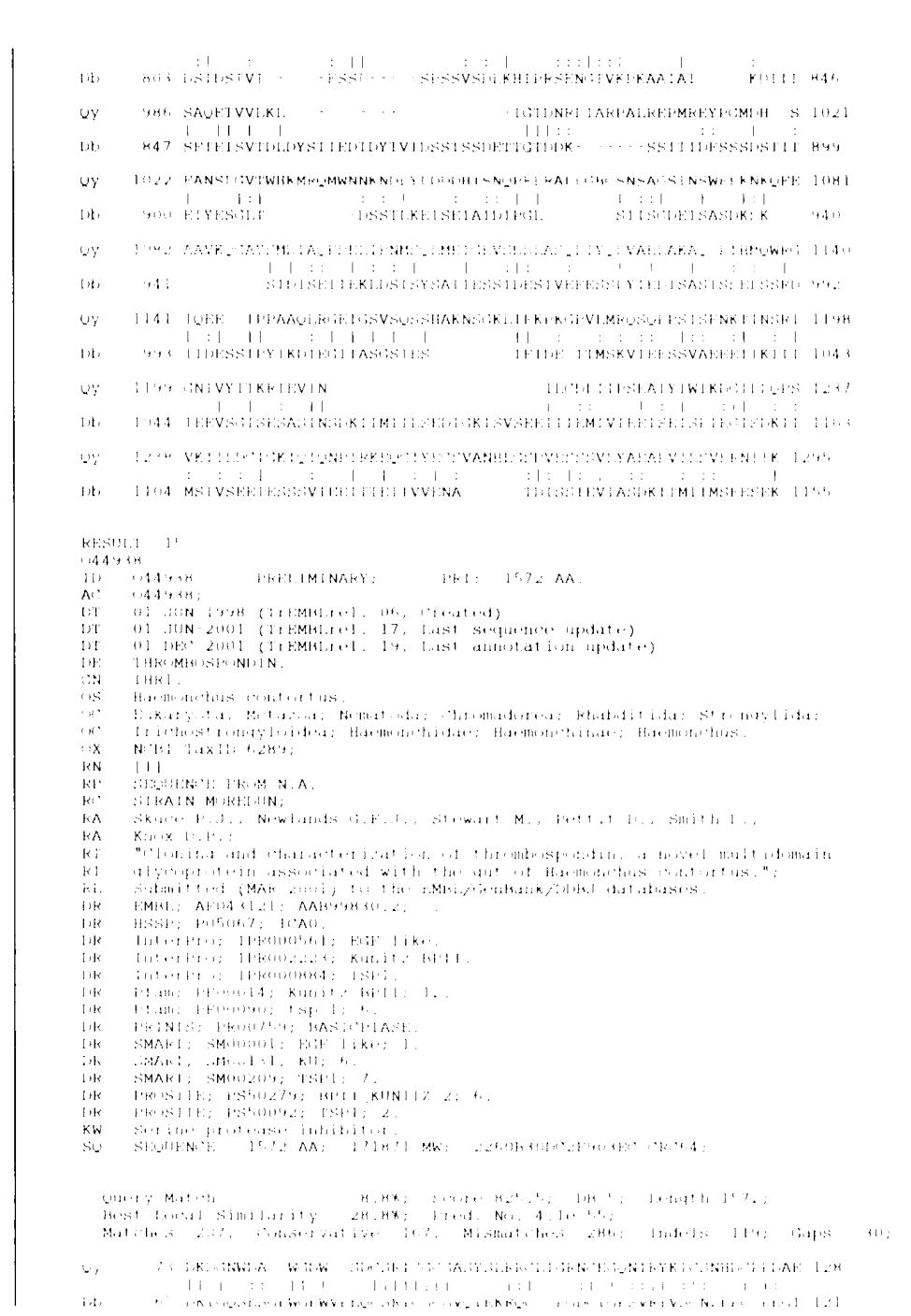
Qy     597 PIQTETELTEEEIQEGPKLFTLPPCLLEACVLPASELLDTPLPEGAEIIYDWEYAGFIPC 656
                  : :|   | |: |||:   | |          |      |:|  : :: |
Db     553 -  --- ----PSQGEAYPLEK CLVDNSTEIPTPTPRSCATLP -  PCEYEWTVSEWSEC 597

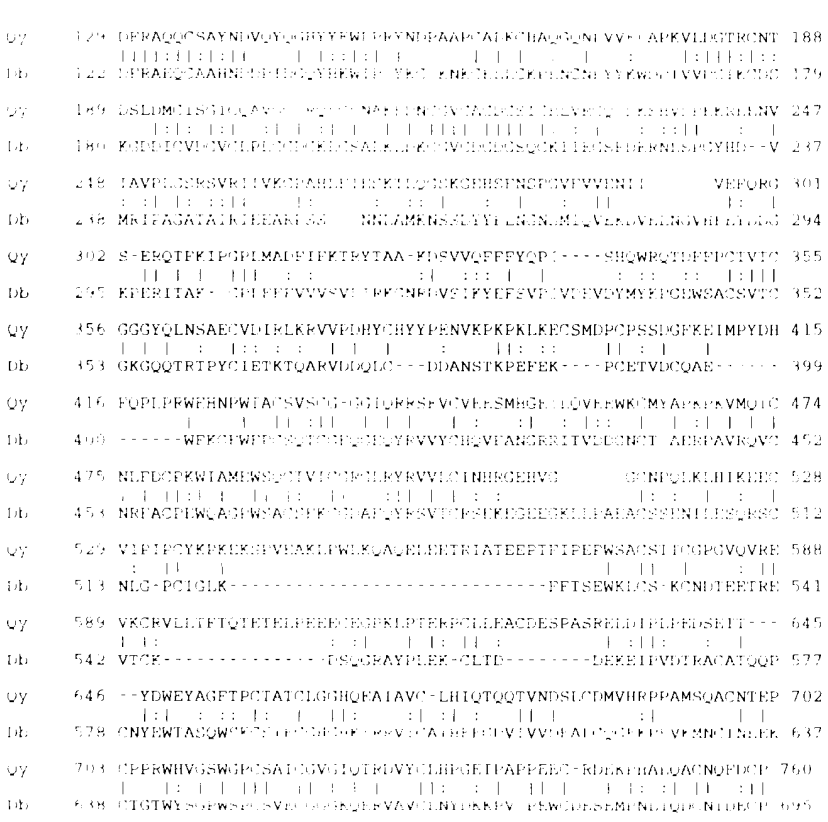
Qy     657 IATCIGGHQEAIAVCLRIQF--QQIVNDSLCTPMVHPPPAMSQACNIEPCPPRWHVGSWGP 714
           |  |  ||:    :|   |    : |:: |   :|   :|   : | | |   |:  ||
Db     598 IIPSGHGHAIEVVIAIHQNSILEVVFPQDT PAIEEGPLNCIGLEFCPPRWV---WGG 656

Qy     715 CSATCGCVCIQTRDVYCLHPCETPAPPEECRDE--KPHALQACNQFDCP 760
           ::| || | | |   ||:  : |  :| | || ||   | ||  |||
Db     657 CIAECGGPSQDRVAVCLNYDKKPV-PEWC-PEAVKPSEKQDCNVDDCP 702


RESULT  14
Q9U8G8
ID   Q9U8G8      PRELIMINARY;      PRT;   4198 AA.
AC   Q9U8G8;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   LACUNIN PRECURSOR.
OS   Manduca sexta (Tobacco hawkmoth) (Tobacco hornworm).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota, Neoptera, Endopterygota, Lepidoptera, Glossata; Ditrysia;
OC   Sphingiodea; Sphingidae; Sphinginae; Manduca.
OX   NCBI_TaxID=7130;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99457716; PubMed=10528409;
RA   Nardi J.B., Martos R., Walden K.K., Lampe D.J., Robertson H.M.;
RT   "Expression of lacunin, a large multidomain extracellular matrix
RT   protein, accompanies morphogenesis of epithelial monolayers in Manduca
RT   sexta.";
RL   Insect Biochem. Mol. Biol. 29:883-897(1999).
DR   EMBL; AF078161; AAF04457.1; -.
DR   HSSP; P02751; 2BNI.
DR   InterPro; IPR004094; Antistasin.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR002223; Kunitz_BPTI.
DR   InterPro; IPR000884; TSP1.
DR   InterPro; IPR002223; WAP.
DR   Pfam; PF02822; Antistasin; 4.
DR   Pfam; PF00047; ig; 2.
DR   Pfam; PF00014; Kunitz_BPTI; 9.
DR   Pfam; PF00095; wap; 1.
DR   PRINTS; PR00759; BASICPTASE.
DR   SMART; SM00408; IGc2; 2.
DR   SMART; SM00131; KU; 10.
DR   SMART; SM00209; TSP1; 7.
DR   SMART; SM00217; WAP; 1.
DR   PROSITE; PS00017; 4_DISULPHIDE_CORE; 1.
DR   PROSITE; PS00280; BPTI_KUNITZ_1; 8.
DR   PROSITE; PS50279; BPTI_KUNITZ_2; 10.
DR   PROSITE; PS50092; TSP1; 1.
KW   Immunoglobulin domain; Serine protease inhibitor; Signal.
```







Search completed: August 19, 2002, 07:53:16
Job time: 515 sec

Gen' ic version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

SM protein - protein search, using sw model

Run on: August 19, 2002, 13:34:54 ; Search time 25.14 Seconds
(without alignments)
2604.406 Million cell updates/sec

Title: US-09-784-358-2
Perfect score: 9446
Sequence: 1 MASWISLPWWVLLRMVEMEGP..... ...SLNDCTIQPYK_R*T_S*_EI 1245

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 48716550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt 40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	9472	100.3	1245	1	ADS9_HUMAN	Q9UKP4 Homo sapien
2	8065.5	85.4	1564	1	AD12_HUMAN	P59437 Homo sapien
3	8186.5	86.8	1077	1	AD19_HUMAN	Q9UKJ8 Homo sapien
4	5993.5	63.4	1205	1	ADS2_BOVIN	P79999 Bos taurus
5	5962	63.1	1291	1	ADS4_HUMAN	O75078 Homo sapien
6	5864.5	62.1	967	1	ADS1_RAT	Q9WTQ1 Rattus norv
7	5771.5	61.1	967	1	ADS1_HUMAN	Q9UKJ9 Homo sapien
8	5643.5	59.8	969	2	AD31_MOUSE	P59671 Mus musculu
9	5641	59.7	1211	1	ADS2_HUMAN	O75676 Homo sapien
10	5443.5	57.6	456	1	AD10_MOUSE	P59438 Mus musculu
11	5209.5	55.2	869	1	AD30_HUMAN	Q9UKP5 Homo sapien
12	4906.5	51.9	890	1	AD38_HUMAN	Q9UQ76 Homo sapien
13	4864.5	51.5	905	2	AD38_MOUSE	P59510 Mus musculu
14	4842	51.2	960	1	AD35_MOUSE	Q9R0C1 Mus musculu
15	4777	50.5	969	1	AD35_HUMAN	Q9UM66 Homo sapien
16	4436	47.0	907	1	AD37_HUMAN	O75173 Homo sapien
17	4151	44.0	867	1	AD34_HUMAN	O95274 Homo sapien
18	4066.5	43.2	630	1	AD34_RAT	Q9QXS9 Rattus norv
19	2923.5	31.0	1522	1	BAI3_HUMAN	O60242 Homo sapien
20	2895.5	30.7	1584	1	BAI1_HUMAN	O14514 Homo sapien
21	276	2.9	903	1	ESP2_XENLA	P54447 Xenopus lae
22	274	2.9	447	1	PK3P_MOUSE	P97478 Mus musculu
23	271	2.9	469	1	PK3P_HUMAN	P27448 Homo sapien
24	265	2.8	807	1	ESP2_RAT	P54446 Rattus norv
25	261.5	2.8	1077	1	SMSA_MOUSE	Q62217 Mus musculu
26	260.5	2.8	1093	2	SMS0_MOUSE	Q99519 Mus musculu
27	257.5	2.7	1034	1	SMSA_HUMAN	O15162 Homo sapien
28	256.5	2.7	8247	1	FGFRM_MOUSE	Q03157 Mus musculu
29	256	2.7	2612	1	ISC7A_HUMAN	O00468 Homo sapien
30	251.5	2.7	1914	1	KMLS_HUMAN	Q15746 Homo sapien
31	251	2.7	1129	1	ISP1_BOVIN	Q28178 Bos taurus
32	246.5	2.7	1522	1	BAI2_HUMAN	O60241 Homo sapien
33	240	2.5	476	1	EB0F_CAVPO	Q14162 Cavia porc
34	240	2.5	1173	1	ISP1_XENLA	P54448 Xenopus lae
35	240.5	2.5	1178	1	ISP2_CHICK	P54446 Gallus gall
36	240.5	2.5	1170	1	ISP2_BOVIN	Q09110 Bos taurus
37	240	2.5	5376	1	ZAN_MOUSE	O88799 Mus musculu
38	240	2.5	1170	1	ISP1_HUMAN	P07996 Homo sapien
39	240	2.5	1172	1	ISP2_HUMAN	P54452 Homo sapien
40	239	2.5	1172	1	ISP2_MOUSE	Q62356 Mus musculu
41	234	2.5	1170	1	ISP1_MOUSE	P54451 Mus musculu
42	233.5	2.5	4393	2	CSPG_HUMAN	P16112 Homo sapien
43	232.5	2.5	837	1	NCM2_MOUSE	Q61136 Mus musculu
44	231.5	2.5	8188	1	BMC1_BOMMO	P08092 Bombyx mori
45	227.5	2.4	1040	1	AXO1_RAT	P20663 Rattus norv

ALIGNMENTS

RESULT 1
ADS9_HUMAN
ID ADS9_HUMAN STANDARD; PRT; 1245 AA.
AC Q9P2N4; Q9NRZ9;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE ADAMTS-9 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 9) (ADAM-TS 9) (ADAM-TS9).
GN ADAMTS9 OR KIAA1312.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (SHORT ISOFORM).
RC TISSUE=Fetal;
RX MEDLINE=20360108; PubMed=11076861;
RA Clark M.E., Kelner G.S., Bartnetzki L.A., Boyer A., Arden K.A.,
RA Maki R.A.;
RT "ADAMTS-9, a novel member of the ADAM-TS/Metallospondin gene
RT family.";
RL Genomics 77:43-50(2000).
RN [2]
RP SEQUENCE OF 1-29 AND 1029-1245 FROM N.A. (LONG ISOFORM).
RC TISSUE=Brain;
RX MEDLINE=20181126; PubMed=11076968;
RA Nagase T., Kikuno R., Ishikawa K.-I., Hirosawa M., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XVI.
RT The complete sequences of 150 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 7:65-73(2000).
CC -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR
CC MATRIX (BY SIMILARITY).
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC SHORT FORM; MAY BE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN ALL FETAL TISSUES.
CC EXPRESSED SLIGHTLY IN ADULT OVARY, PANCREAS, HEART, KIDNEY, LUNG,
CC PLACENTA. ALSO DETECTED IN SPINAL CORD AND BRAIN. NOT DETECTED IN
CC COLON, SMALL INTESTINE, TESTIS, LIVER, SKELETAL MUSCLE, SPLEEN OR
CC THYMUS.
CC -!- DOMAIN: THE SPACER DOMAIN AND THE 1ST TYPE I DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX (BY
CC SIMILARITY).
CC -!- PTM: THE PROTEIN IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY
CC SIMILARITY).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -!- SIMILARITY: CONTAINS 11 TSP TYPE 1 DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```
CC   entities requires a license agreement (See http://www.isb-sib.ch/
CC   or send an email to license@isb-sib.ch).
CC   -!- ------------------------------------------------------------
DR   EMBL; AF262618; AAF91164.1; -.
DR   EMBL; AB037733; BAA94255.1; -.
DR   MIM; 605421; -.
DR   InterPro; IPR001762; Disintegrin.
DR   InterPro; IPR002870; Pep_M12B_propep.
DR   InterPro; IPR001590; Reprolysin.
DR   InterPro; IPR000884; TSP1.
DR   InterPro; IPR000130; Zn_MTpeptdse.
DR   Pfam; PF01562; Pep_M12B_propep; 1.
DR   Pfam; PF01421; Reprolysin; 1.
DR   Pfam; PF00090; tsp_1; 11.
DR   SMART; SM00209; TSP1; 12.
DR   PROSITE; PS50215; ADAM_MEPRO; 1.
DR   PROSITE; PS00427; DISINTEGRIN_1, FALSE_NEG.
DR   PROSITE; PS50092; TSP1; 6.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Hydrolase; Metalloprotease; Signal; Glycoprotein; Zymogen;
KW   Repeat; Extracellular matrix; Alternative splicing.
FT   SIGNAL       1     18       POTENTIAL.
FT   PROPEP      19    287       BY SIMILARITY.
FT   CHAIN      288   1629       ADAM15.
FT   DOMAIN     509    587       DISINTEGRIN-LIKE.
FT   DOMAIN     589    642       CYS-RICH (TYPE I).
FT   DOMAIN     645    752       CYS-RICH.
FT   DOMAIN     753    880       SPACER.
FT   DOMAIN     999   1053       TSP TYPE-1 2.
FT   DOMAIN    1056   1108       TSP TYPE 1 3.
FT   DOMAIN    1111   1156       TSP TYPE-1 4.
FT   DOMAIN    1184   1239       TSP TYPE-1 5.
FT   DOMAIN    1240   1295       TSP TYPE-1 6.
FT   DOMAIN    1342   1383       TSP TYPE 1 7.
FT   DOMAIN    1386   1439       TSP TYPE-1 8.
FT   DOMAIN    1445   1498       TSP TYPE-1 9.
FT   DOMAIN    1501   1554       TSP TYPE-1 10.
FT   DOMAIN    1562   1612       TSP TYPE-1 11.
FT   DOMAIN      88     96       POLY-SER.
FT   SITE       223    223       CYSTEINE SWITCH (POTENTIAL).
FT   METAL      434    434       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE   435    435       BY SIMILARITY.
FT   METAL      438    438       ZINC (CATALYTIC) (BY SIMILARITY).
FT   METAL      444    444       ZINC (CATALYTIC) (BY SIMILARITY).
FT   CARBOHYD   112    112       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   135    135       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   271    271       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   749    749       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   840    840       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD  1217   1217       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD  1267   1267       N-LINKED (GLCNAC...) (POTENTIAL).
FT   VARSPLIC  1361   1372       GLVTGGKGH    VEWEGTTEI (IN
FT                              SHORT ISOFORM).
FT   VARSPLIC  1073   1629       MISSING (IN SHORT ISOFORM).
FT   CONFLICT   367    367       E -> L (IN REF. 1).
SQ   SEQUENCE  1629 AA;  182649 MW;  C124CEFF5B8B941F CRC64;


  Query Match           10.4%;  Score 972;  DB 1;  Length 1629;
  Best Local Similarity  14.3%;  Pred. No. 1e-53;
  Matches  326;  Conservative  145;  Mismatches  405;  Indels  814;  Gaps   56;

Qy      75 DGNWDAWCDWSDCSRFCGGGASYSLRKCLT-----GRNCEGQNIRYKFCNMHECPPDAF 128
           ||:| :|  :  |||||||||  :|| |       :| | |: :::|:|:   |
Db     548 DGSWGSWSPPGPCSRPCGGGIRTASRFCNSPEPANGGKCVGSDVQYQSCNTEACPVDGS 607

Qy     129 DFRAQQCNGAYNCVQYQHYY          FWFDAYNCTRAAF  CAIMGHAGQGNLVVILA 177
           ||| :||:       : | ::           |:|:|:       | | |   |     :|
Db     608 DFRDEQCKA     RPCDRDVCNGRGICNNVSWVFEYDTIMECKCEIDCRKVACNFAYVQLK 702

Qy     178 PEVEVSEPMNITSSLEMTSGIQVAVDTRGDSNARLAVNTGVDVRKRSV  236
           :|||| |  |: |:::  :||  : | | | | |||| || |:|: | |   :
Db     703 DRVIDGIPCGQDTNGICVQGECNGAGICHVLNGKARECTGAVDGRDVVAMGIGV 756
```

```
Qy     237 HVSPSPRREPNVIAVPLCGPSVRITVPGPAHLTIPCR   - -TLQTSKGPHSFNSPGVFV 290
           |      |: :| |: : :  :    | |        |  ||||   |  | ||
Db     757 HYG    YNTVVKTFAGATNIDVR    QHSPSGHTDDDNYLALSSGKGEFLLN  GNFV 812

Qy     291 VE--       -NTTVEPQKGSE       PQTPFIPGPLMAPPIP    - -KIRYT-- A 325
           |          | ||:  |||         | :|   |:   :           ||:
Db     813 VTMAKREIRIGNAVVEYS GSETAVERINSTDRIEQELLLQVLSVGKLYNPDVRYSFNIP 871

Qy     326 AKDSVVQFFF-----YQPISH------------------------------------ 341
           :|   |||:      :|  |
Db     872 IEDKPQQFYWNSHGPWQACSKPCQGERRRKLVCTPFSDQLTVSPQPCDPLPQPGHITEPC 931

Qy     342 ------QWPQTDFFPCTVTCGSGYQLNSAEC-----VDIRLKRVVPDHYCHYYPENVKPK 390
                :|       |:  || ||:      :     :| : :: | | :|  :      ||
Db     932 GTDCQLWHVASSSCPATCDSYRIIKYCAFYSRTPTPTF VYGPTPSSP     -PP 985

Qy     391 PKLPECGMETCFSSGRGIKEIMFYDGPQFLEFEWEHNFWTACSVTCGGGIQERSFVCVEESM 450
           | :|     | :                  | :: || || || || ||| :|| :
Db     986 PSNPFA TVRLCNIGG      --          WEYDAWIEDPPC  -  TLPSKKAICV  NT 1027

Qy     451 PGDWICPDCPESGRNDTQMYFGLCAAVGSILTFIVIALLVLVMYRMKKQDYGSIDLHLNRG 510
           :|  :: || :  |  : | | | ||| : :|:  | |  |  |||| : :  | |
Db    1028 ENDVL  HCDFCHGGRVCSSKGVCYCEDGWEGPDCSVPVQHPHSGQVWCQEGED 1084

Qy     511 INNNSGMAVWGDVNRDLLESLPLVSLATVPIPTEGTVEAGSWFEVGAEIKPSCLQHCEAIE 570
           :| |      ||: |    : |                                  |
Db    1085 RLNDR-----MCDPETKPISMQTCQQP --- -     ---      -----    --E 1107

Qy     571 EPTPIPEPWSACSTTCGPGVQVREVKCRVLLIPIQTETELPEEECEGPKLPTE RPCLLE 630
           ::  ||  || ||| | |:| ||| :: |:      : : :|     || : | :
Db    1108 CASWQAGPWGQCSVICGQGYQLRAVKC-IIGTYMSV---VDDNDCNAATRPTDTQDCELP 1163

Qy     631 ACDESPASRELDIPLPEDSETTY-----DWEYAGFIPCTAPCIGGHQEATAVCLHIQTQQ 685
           :|  ||:      ||   :||     | :  :|||:|||  | :      |  :
Db    1164 SCHPPPAA-------PEIPPSIYSAPPIQWPEPCSWIPCSATCGKGTRMRYVSC-EDENGS 1215

Qy     686 TVNDSLCDMVHRPPAMSQACNTEPCFPRWHVGSWGPCSATCGVGTQTRDVYCLIPGEIVA 745
           :: | |  : || |  : |:  ||  :|     |  || ||| |  || | |::  :
Db    1216 VADLSACATLPKPVA KLPCSVTYC GQWKAPDWSGCNVICGQGVRQRQVMCVRYSDHVI 1273

Qy     746 PPEEC-RDEKPHALQACNQFDCPP------------------ ------------GWH 784
           || :|  |   || :   ||                                     |
Db    1274 IPSP-TLGAKIPEPVQACMGRPCPLPIPGAPAGPEGNEDERPKSASPSKIHVIGGNQWR 1333

Qy     785 TEEWQQCSATCGSGIQERVVDCQQTFDGSPASDSAYCHAEDRPESRQCNTDPCPIDGAW 844
           |  || || ||:| | | |:    :|   |  :  :  |    ::|    | |  |
Db    1334 TGEWCACSPTCAGGIQERVVVCQQ  LNGTPAN   PGVLCIKEVEGQGNACESGPC PQWA 1387

Qy     845 VGDWSPCSVTCGPGIRERSVECAAPCDGRPVLQRRMCNLHPCPIDGAWGPWSPCSVTCGKG 904
           |:| :|:  || ||: |  ||||  : | | |        ||        ||:
Db    1388 YGNWGLCTKTCGGGTKTRTVVCQQK  GNGEKEF       DG       GCL      - 1423

Qy     905 SEMETRICPQGPQIIQVCEVYIQTPELEKRINLTIGSEAYLLPNISVIIKCPVPPPQKSLI 964

Db    1424 ------------------------------------ - ----   -----      --  1423

Qy     965 QWEKDGPCLQNSKRIGIIKSGSLKIHGLAAPDIGVYRCIAGSAQETVVLKLIGTDNRLIA 1024

Db    1424 -------------                                                 1423

Qy    1025 PPAIEFPMPEYPQMGHSPANSIGVTWHKMPQMWNNVNDLYLPDGHTSNQPFLPAIIGPGS 1084

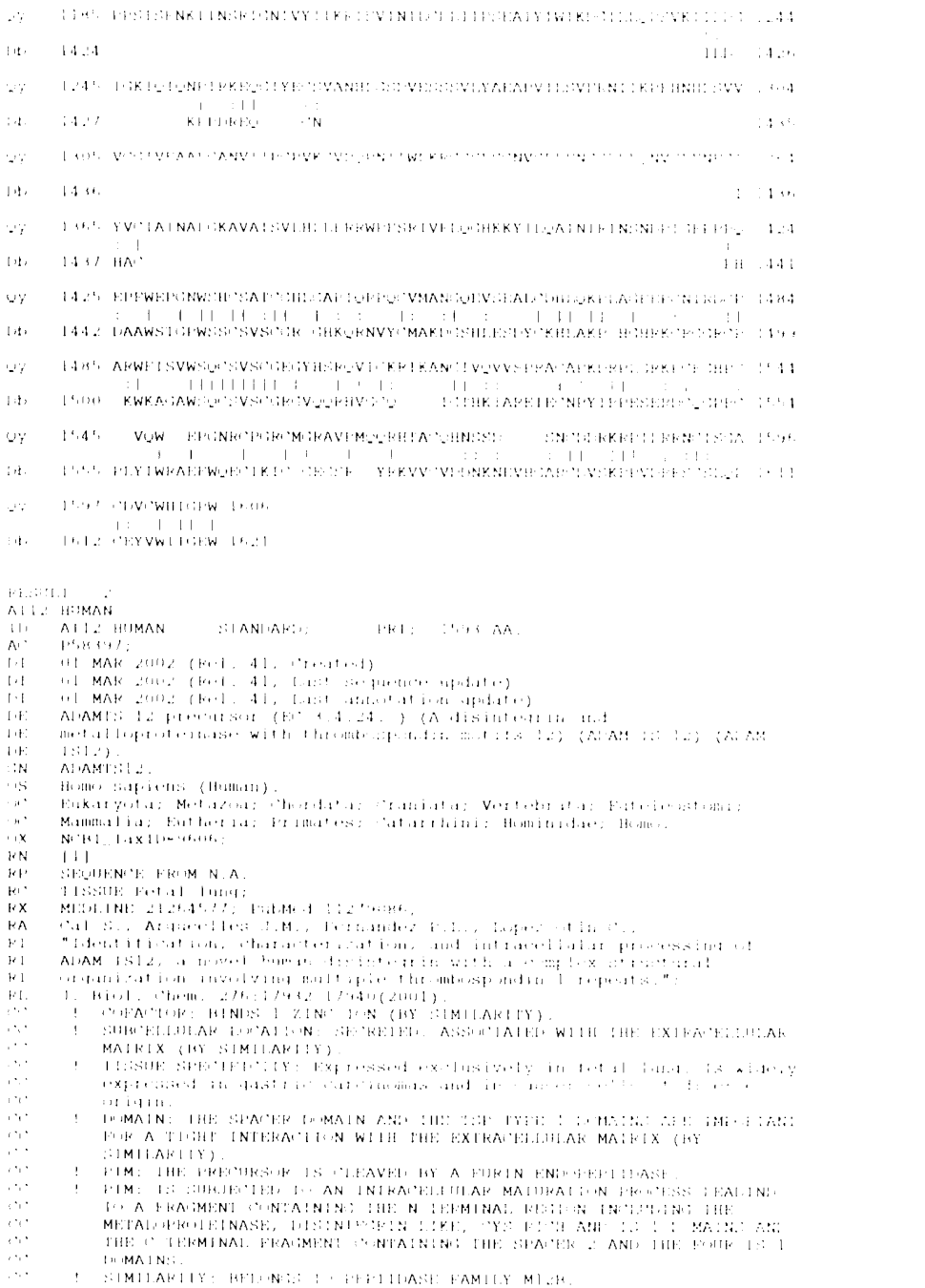
Db    1424 ---------------  ------------------------------------------ ----- 1423

Qy    1085 NSAGSTNNSWELPNRQLFAAVPLQAYSMSTAQFEKLTRMMSQLMKLSTVSKNLKDYQL 1144

Db    1424 ---------------    ---        ----   ------    -  ---      ---- 1423

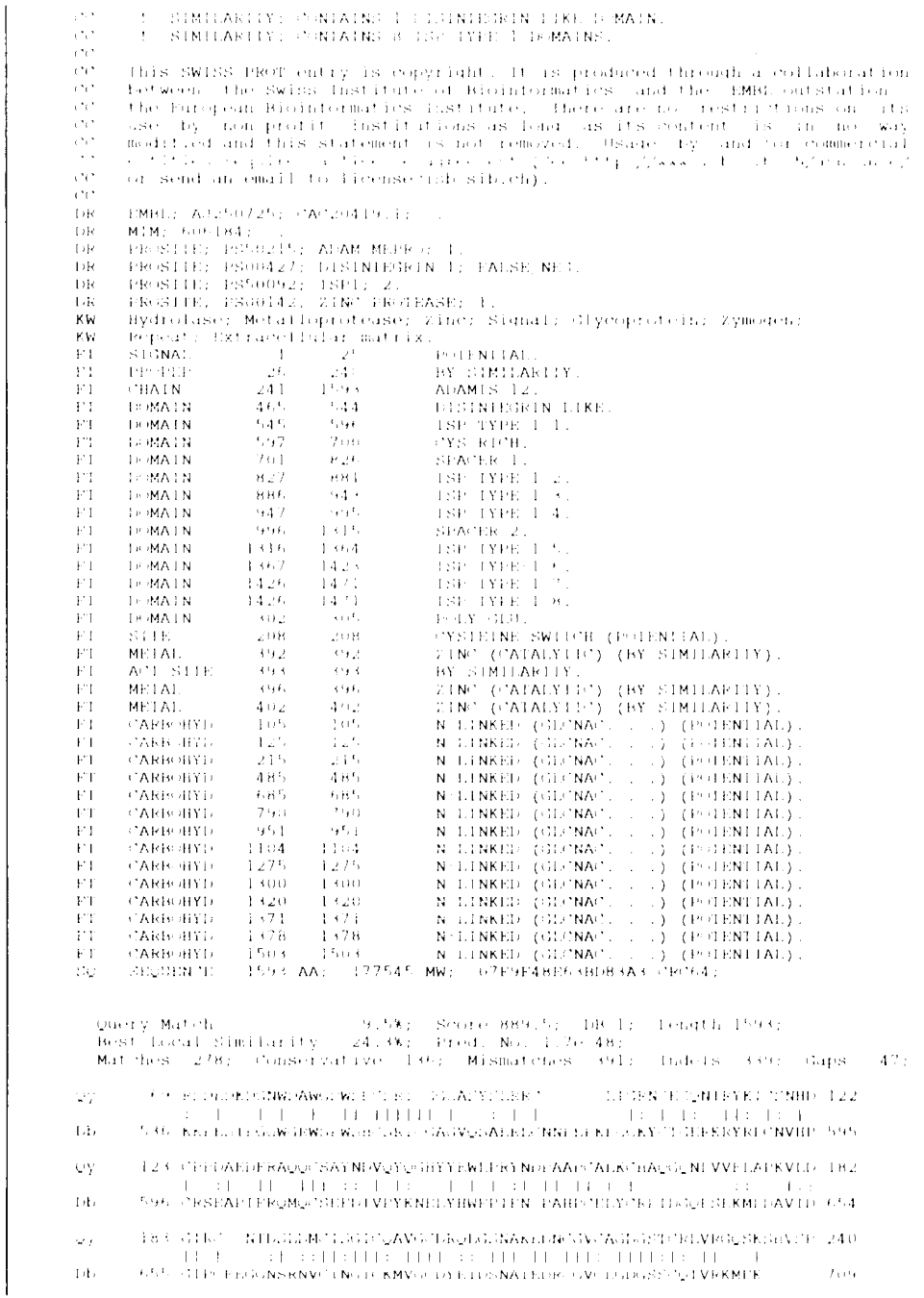
Qy    1145 VAELAKAQFHMQWSQLQEFIDAAQLEMTIGSVSQSHAKNSCKLLKFPEGHVLMPQSQ 1204

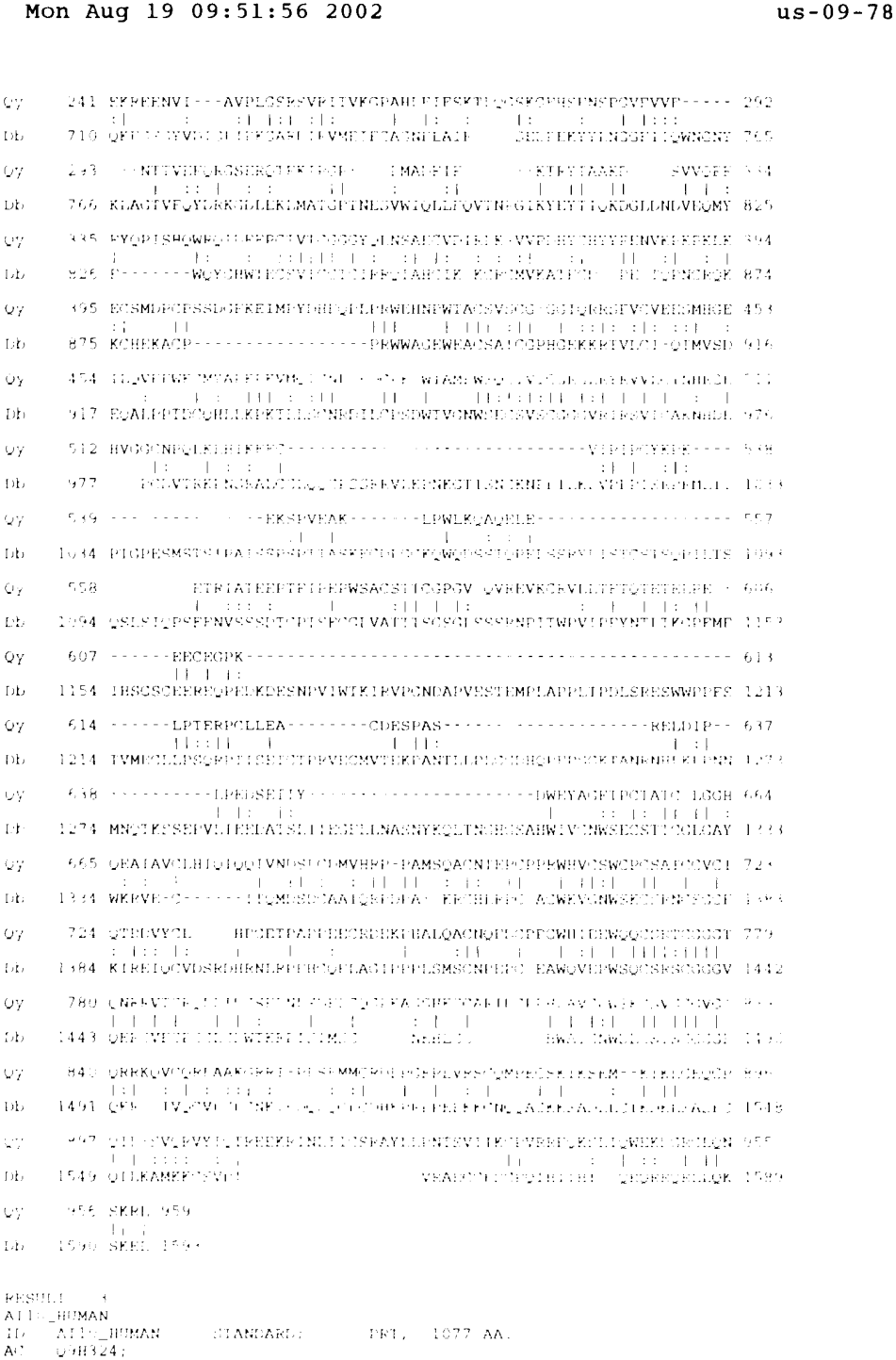
Db    1424 ----     --------- -------- ------ - ------ -  ----      --- 1423
```



```
ID   ADA12_HUMAN     STANDARD;      PRT;   1503 AA.
AC   P78325;
DT   01-MAR-2002 (Rel. 41, Created)
DT   01-MAR-2002 (Rel. 41, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   ADAMTS 12 precursor (EC 3.4.24.-) (A disintegrin and
DE   metalloproteinase with thrombospondin motifs 12) (ADAM-TS 12) (ADAM
DE   TS12).
GN   ADAMTS12.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Fetal lung;
RX   MEDLINE=21264577; PubMed=11278096;
RA   Cal S., Arguelles J.M., Fernandez P.L., Lopez-Otin C.;
RT   "Identification, characterization, and intracellular processing of
RT   ADAM-TS12, a novel human disintegrin with a complex structural
RT   organization involving multiple thrombospondin-1 repeats.";
RL   J. Biol. Chem. 276:17932-17940(2001).
CC   -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR
CC       MATRIX (BY SIMILARITY).
CC   -!- TISSUE SPECIFICITY: Expressed exclusively in fetal lung. Is widely
CC       expressed in adult connective tissues and in tissues with a high
CC       origin.
CC   -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE 1 DOMAINS ARE IMPORTANT
CC       FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX (BY
CC       SIMILARITY).
CC   -!- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE.
CC   -!- PTM: IS SUBJECTED TO AN INTRACELLULAR MATURATION PROCESS LEADING
CC       TO A FRAGMENT CONTAINING THE N-TERMINAL PROTEIN INCLUDING THE
CC       METALOPROTEINASE, DISINTEGRIN LIKE, CYS-RICH AND TSP-1 MAJOR AND
CC       THE C-TERMINAL FRAGMENT CONTAINING THE SPACER 2 AND THE FOUR TSP-1
CC       DOMAINS.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC   -!- SIMILARITY: CONTAINS 1 DISINTEGRIN LIKE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 4 TSP TYPE 1 DOMAINS.
CC
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC
DR   EMBL; AJ250725; CAC20419.1; -.
DR   MIM; 606184; -.
DR   PROSITE; PS50215; ADAM_MEPRO; 1.
DR   PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
DR   PROSITE; PS50092; TSP1; 2.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW   Repeat; Extracellular matrix.
FT   SIGNAL       1     21       POTENTIAL.
FT   PROPEP      22    214       BY SIMILARITY.
FT   CHAIN       215   1503      ADAMTS 12.
FT   DOMAIN      465    544      DISINTEGRIN LIKE.
FT   DOMAIN      545    598      TSP TYPE 1 1.
FT   DOMAIN      597    709      CYS RICH.
FT   DOMAIN      701    826      SPACER 1.
FT   DOMAIN      827    884      TSP TYPE 1 2.
FT   DOMAIN      886    943      TSP TYPE 1 3.
FT   DOMAIN      947    995      TSP TYPE 1 4.
FT   DOMAIN      996   1315      SPACER 2.
FT   DOMAIN     1316   1364      TSP TYPE 1 5.
FT   DOMAIN     1367   1425      TSP TYPE 1 6.
FT   DOMAIN     1426   1477      TSP TYPE 1 7.
FT   DOMAIN     1426   1471      TSP TYPE 1 8.
FT   DOMAIN      502    505      POLY GLU.
FT   SITE        208    208      CYSTEINE SWITCH (POTENTIAL).
FT   METAL       392    392      ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    393    393      BY SIMILARITY.
FT   METAL       396    396      ZINC (CATALYTIC) (BY SIMILARITY).
FT   METAL       402    402      ZINC (CATALYTIC) (BY SIMILARITY).
FT   CARBOHYD    105    105      N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    125    125      N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    215    215      N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    486    486      N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    665    665      N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    798    798      N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    951    951      N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   1104   1104      N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   1275   1275      N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   1300   1300      N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   1420   1420      N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   1471   1471      N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   1478   1478      N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   1503   1503      N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   1503 AA;  172545 MW;  67F9F48E6BD84A3 CRC64;
```

```
  Query Match          9.5%;  Score 889.5;  DB 1;  Length 1503;
  Best Local Similarity   24.4%;  Pred. No. 1.2e-48;
  Matches  278;  Conservative  146;  Mismatches  391;  Indels  330;  Gaps  47;
```





```
RESULT  4
AT1C_HUMAN
ID   AT1C_HUMAN     STANDARD;      PRT,  1077 AA.
AC   Q9H324;
```

```
DT   01-MAR-2002 (Rel. 41, Created)
DT   01-MAR-2002 (Rel. 41, Last sequence update)
DT   31-MAR-2002 (Rel. 41, Last annotation update)
DE   ADAMTS 10 precursor (EC 3.4.24. ) (A disintegrin and metalloproteinase
DE   with thrombospondin motifs 10; ADAM-TS 10) (ADAM-TS10) (Fragment).
GN   ADAMTS10.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Apte S.S.;
RT   "ADAM-TS10: a novel member of the ADAM-TS family containing multiple
RT   thrombospondin type 1 repeats.";
RL   Submitted (JUN 1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- CAUTION: BLAST 1 LINE ION (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR
CC       MATRIX (BY SIMILARITY).
CC   -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE 1 DOMAINS ARE IMPORTANT
CC       FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX (By
CC       similarity).
CC   -!- SIMILARITY: BELONGS TO THE ADAMTS FAMILY OF M12B.
CC   -!- SIMILARITY: CONTAINS 1 DISINTEGRIN LIKE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 5 TSP TYPE-1 DOMAINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AF163762; AAG35563.1; -.
DR   InterPro; IPR002870; Pep_M12B_propep.
DR   InterPro; IPR001590; Reprolysin.
DR   InterPro; IPR000884; TSP1.
DR   InterPro; IPR000179; Zn_MIpeptdse.
DR   Pfam; PF01562; Pep_M12B_propep; 1.
DR   Pfam; PF01421; Reprolysin; 1.
DR   Pfam; PF00090; tsp_1; 5.
DR   SMART; SM00209; TSP1; 5.
DR   PROSITE; PS50215; ADAM_MEPRO; 1.
DR   PROSITE; PS50092; TSP1; 2.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
DR   PROSITE; PS00127; MSTNPRO_1; FALSE_NEG.
KW   Hydrolase; Metalloprotease; Zinc; Glycoprotein; Zymogen;
KW   Repeat; Extracellular matrix.
FT   NON_TER       1      1
FT   PROPEP        <1     207       BY SIMILARITY.
FT   CHAIN         208    1077      ADAMTS 10.
FT   METAL         365    367       ...IN (...IALYTIC) (BY SIMILARITY).
FT   ACT_SITE      367    367       BY SIMILARITY.
FT   METAL         371    373       ZINC (CATALYTIC) (BY SIMILARITY).
FT   METAL         376    376       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DOMAIN        424    520       DISINTEGRIN-LIKE.
FT   DOMAIN        578    679       CYS-RICH.
FT   DOMAIN        680    862       SPACER.
FT   DOMAIN        521    577       TSP TYPE-1 1.
FT   DOMAIN        799    860       TSP TYPE-1 2.
FT   DOMAIN        862    918       TSP TYPE-1 3.
FT   DOMAIN        922    976       TSP TYPE-1 4.
FT   DOMAIN        981    1031      TSP TYPE-1 5.
FT   CARBOHYD       64     64       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD      196    196       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD      297    297       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD      714    714       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD      769    769       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD      866    866       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1077 AA;  118072 MW;  3914DE18DCBBF587 CRC64;
```

```
  Query Match              4.8%;  Score 818.5;  DB 1;  Length 1077;
  Best Local Similarity    26.7%;  Pred. No. 3.2e-44;
  Matches  208;  Conservative  80;  Mismatches  194;  Indels  291;  Gaps  26;

Qy         72 EDKIGNWDAWGDWSDCSRTCGGASYSLRRC       LTGRNCEGQNIRYKTCSNHDCPP 125
Db        518 EGVDGAWGPWTPWGDCSRTCGGGVSSSRRHCDSPRPTIGGKYCLGERRRHRSCNTDDCPP 577

Qy        126 DAEDFRAQQCSAYNDVQYQGHYYEWLPRYNDPAAPCALKCHAQGQNLVVELAPKVLDGTR 185
Db        578 GSQDFREVQCSEFDSIPFRGKFYKWKTYRGGGVKACSLTCLAEGFNFYTERAAAVVDGTP 637

Qy        186 CNTDSLDMCISGICQAVGCDRQLGSNAKEDNCGVCAGDGSTCRLVRGQSKSHVSPEKREE 245
Db        638 CRPDTVDICVSGECKHVGCDRVLGSDLREDKCRVCGGDGSACETIEGVFSPASPGAGYE 696

Qy        246 NVIAVPLGSRSVRITVKGPAHLFI               ESKTLQGSKGEHSFNSPGVF 289
Db        697 DVVWIP            KGSVHIFIQDLNLSLSHLALKGDQESLLLEGLPGT   PQPHRL 743

Qy        290 VVENTTVEFQRGSERQTFKIPGPLMADFIF     KTRYTAAKDSVVQFFFYQPIS     340
Db        744 PLAGTTFQLRQGPDQVQSLEALGPINASLIVMVLARTELPA       LRYRFNAPIARDSL 798

Qy        341    HQWRQTDFFPCTVTCGGGYQLNSAECVDIRLKRVVPDHYCHYYPENVKPKPKLKECSM 398
Db        799 PPYSWHYAPWTKCSAQCAGGSQVQAVECRNQLDSSAVAPHYCSAHSK   LPK RQRACNT 855

Qy        399 DPCPSSDGFKEIMPYDHFQPLPRWEHNPWTACSVSCGGGIQRRSFVCVEESMHGEILQVE 458
Db        856 EPCP                      PDWVVGNWSLCSRSCDAGVRSRSVVCQRRVSAAEEKALD 898

Qy        459 EWKCMYAPKPKVMQTCNLFDCPKWIAMEWSQCTVTCGRGLRYRVVLCINHRGEHVGGCNP 515
Db        899 DSACPQPRPPVLEACHGPTCPPEWAALDWSECTPSCGPGLRHRVVLCKSADHR        951

Qy        516 QNPQEKRHEKTPEPVTPTPTYKPKEKSPVHAKTPWLKQALPLEEPTPLATEEPTEEPAQS 575
Db        952                                                             951

Qy        576 CSTPSGPGVQVREVKCRVLLPTTQTETTSLPEEEERSPKEPTEPPCLLEACTESPACPEL 635
Db        952                               ATLPPAHC                  SPAA   963

Qy        636 IPLPEDSETTYDWEYAGFTPCTATCLGGHQEATAVCLHTQTQQTVNDSLCDMVHRPPAMS 695
Db        964                                                   KFPA C 968

Qy        696 QACNTEPCPP RWHVGSWGPCSATCGVGIQTRDVYCL HPG                   734
Db        969 MRCNLRRCPPARWVAGEWGECSAQCGVGQRQRSVRCTSHTGQASHECTEALRPPTTQQCE 1028

Qy        735      ETPAP    PEECRPEKTHAQQACNQTPCTGWITEEWQQCS           PTCGG 777
Db       1029 AKCDSPTPGDGPEECKDVNKVAY       CP    LVLKFQFCSRAYFRQMCCKTCQG   1076


RESULT 4
AIS2_BOVIN
ID   AIS2_BOVIN     STANDARD;      PRT;  1205 AA.
AC   P79331;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   ADAMTS 2 precursor (EC 3.4.24.14) (A disintegrin and
DE   metalloproteinase with thrombospondin motifs 2) (ADAM-TS 2) (ADAM-TS2)
DE   (Procollagen I/II amino propeptide processing enzyme) (Procollagen I
DE   N-proteinase) (PC I-NP) (Procollagen N-endopeptidase) (pNPI).
GN   ADAMTS2 OR NPI.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Skin;
RX   MEDLINE=97225960; PubMed=9122202;
RA   Colige A., Li S.W., Sieron A.L., Nusgens B.V., Prockop D.J.,
RA   Lapiere C.M.;
RT   "cDNA cloning and expression of bovine procollagen I N-proteinase: a
RT   new member of the superfamily of zinc-metalloproteinases with binding
RT   sites for cells and other matrix components.";
RL   Proc. Natl. Acad. Sci. U.S.A. 94:2374-2379(1997).
RN   [2]
RP   PARTIAL SEQUENCE.
RX   MEDLINE=95348096; PubMed=7622483;
RA   Colige A., Beschin A., Samyn B., Goebels Y., Van Beeumen J.,
RA   Nusgens B.V., Lapiere C.M.;
RT   "Characterization and partial amino acid sequencing of a 107-kDa
RT   procollagen I N-proteinase purified by affinity chromatography on
RT   immobilized type XIV collagen.";
RL   J. Biol. Chem. 270:16724-16730(1995).
CC   -!- FUNCTION: CLEAVES THE PROPEPTIDES OF TYPE I AND II COLLAGEN PRIOR
CC       TO FIBRIL ASSEMBLY. DOES NOT ACT ON TYPE III COLLAGEN. MAY ALSO
CC       PLAY A ROLE IN DEVELOPMENT THAT IS INDEPENDANT OF ITS ROLE IN
CC       COLLAGEN BIOSYNTHESIS.
CC   -!- CATALYTIC ACTIVITY: Cleaves the N-propeptide of collagen chain
CC       alpha 1(I) at Pro-|-Gln and of alpha 1(II) and alpha 2(I) chains
CC       at Ala-|-Gln.
CC   -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC   -!- SUBUNIT: MAY BELONG TO A MULTIMERIC COMPLEX. BINDS SPECIFICALLY TO
CC       COLLAGEN TYPE XIV.
CC   -!- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR
CC       MATRIX (BY SIMILARITY).
CC   -!- TISSUE SPECIFICITY: ENZYMATIC ACTIVITY IS DETECTED AT HIGH LEVEL
CC       IN ALL TYPE I COLLAGEN RICH TISSUES SUCH AS SKIN, BONES, TENDONS
CC       AND AORTA AND AT LOW LEVEL IN BRAIN AND THYMUS. THE MRNA LEVELS
CC       WERE DISPROPORTIONATELY HIGH IN HEART, LIVER, RETINA AND MUSCLE.
CC   -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
CC       FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC   -!- PTM: THE N-TERMINUS IS BLOCKED.
CC   -!- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY
CC       SIMILARITY).
CC   -!- DISEASE: DEFECTS IN ADAMTS2 ARE THE CAUSE OF DERMATOSPARAXIS, A
CC       RECESSIVELY INHERITED DISORDER CHARACTERIZED BY SEVERE SKIN
CC       FRAGILITY AND BIOCHEMICALLY BY THE PRESENCE IN SKIN OF PROCOLLAGEN
CC       INCOMPLETELY PROCESSED AT THE AMINO TERMINUS.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC   -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 4 TSP TYPE-1 DOMAINS.
CC   -!- CAUTION: HAS SOMETIMES BEEN REFERRED TO ADAMTS-3.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X96389; CAA65253.1; -.
DR   InterPro; IPR001762; Disintegrin.
DR   InterPro; IPR002870; Pep_M12B_propep.
DR   InterPro; IPR001590; Reprolysin.
DR   InterPro; IPR000884; TSP1.
DR   InterPro; IPR000130; Zn_MIpeptdse.
DR   Pfam; PF01562; Pep_M12B_propep; 1.
DR   Pfam; PF01421; Reprolysin; 1.
DR   Pfam; PF00090; tsp_1; 4.
DR   SMART; SM00209; TSP1; 4.
DR   PROSITE; PS50215; ADAM_MEPRO; 1.
DR   PROSITE; PS50092; TSP1; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; FALSE_NEG.
DR   PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
KW   Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
```

```
KW   Repeat; Collagen degradation; Extracellular matrix; Heparin-binding.
FT   SIGNAL        1     28       POTENTIAL.
FT   PROPEP       29    253       BY SIMILARITY.
FT   CHAIN       254   1205       ADAMTS 2.
FT   METAL       402    402       ZINC (CATALYTIC) (POTENTIAL).
FT   ACT_SITE    403    403       BY SIMILARITY
FT   METAL       406    406       ZINC (CATALYTIC) (BY SIMILARITY).
FT   METAL       412    412       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DOMAIN      474    554       DISINTEGRIN-LIKE.
FT   DOMAIN      555    611       TSP TYPE-1 1.
FT   DOMAIN      612    736       CYS-RICH.
FT   DOMAIN      737    845       SPACER.
FT   DOMAIN      846    905       TSP TYPE 1 2.
FT   DOMAIN      906    966       TSP TYPE 1 3.
FT   DOMAIN      969   1024       TSP TYPE 1 4.
FT   SITE        685    687       CELL ATTACHMENT SITE (POTENTIAL).
FT   DOMAIN       41     85       POLY-ALA.
FT   DOMAIN      177    190       POLY-GLU.
FT   CARBOHYD    184    184       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    245    245       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    942    942       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    943    943       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    987    987       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1025   1025       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1052   1052       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1139   1139       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1144   1144       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1205 AA;  133887 MW;  7B7B232A15320271 CRC64;


  Query Match             6.4%;  Score 593.5;  DB 1;  Length 1205;
  Best Local Similarity   27.9%;  Pred. No. 7.4e-30;
  Matches  178;  Conservative   86;  Mismatches  266;  Indels  109;  Gaps   27;

Qy      74 KDGNWDAWCTDWSPCSKTCGGGAQYSLRFCL-------GSNTGCVGNKSKTCSNHDCPTDVA 127
           :|||| ||   :  ||||.|| |   :  |:|       || | |    :: |:: |||
Db     554 RDGNWGAWSPEGSCSRTCGGGVKFRYRQCDNPRPANGGKYCVGSGLAYDFQLCNSQDCPDAL 613

Qy     128 KDFRAQQCSAYN---   DVQVLCHYYEWLP RYNDPAADCALRCHAQCQNLVVFLAEEV 180
           ||: :||  ::      | |     : ||| : |    | | . ::    || :   |
Db     614 ADFREDGCWQWQWPLYYFPRCIAQ---HHWLPHFHRDAKLRCHLYCTSKFTGNVVSMKRMV 669

Qy     181 LDGTRCN TGKICMSGKCVDLGAV-TKKQLGSNAKLDWISVFGIDLSLNTHYLFVANQGK 239
           |||||:  |:  ::: | |:|||| :||: :|| | || || | |:::|:|     |
Db     669 HDCIKDLVSKGAVSVTGVTSQSWNCRCYATIDFNRDAVVKSLFGE  K 716

Qy     240 PEKPPP-NVTAVPQNPVPETVKPTPAHLTVSKPTTQPP   QPHSPNCPS  VPVVPN 293
           |:|     :| | :|  :  |          :  | |:  |    | :   |    |:
Db     717 PEKDGYIEMPELIADAVRILDFEAKVTGIHLLAVFVKLSTIYGQTIKRSLFKMAV 770

Qy     294 IIVEPQRGSEG_IPKTPCPEM-           AGTIPPTETTAAKLVV   VQPTPP_ 337
           |::    |:| :  |||          |     :|   :||:
Db     771 VEWEYKDHRGRPRTLQIMGPLHGLIIVELVIPEGDARISLIYKYMIHEDSLNVIRNNVLEDD 830

Qy     338 PISHLWELIFTTIVTFTYLLNSAFCVCIFLKFVSLVRLRVAS-RTKEDV  385
           : :::    : |||:  |||| |     |      :::|   :|    :::| |||   : |
Db     831 SVGIEWALKEWSPCSKPCGGGSQFRYRMCKRRDGMVHPGSCD--PDSVSKPKAIPPDC 888

Qy     386 SMDPCPSSIQFKKEIMPYDHFQPLPFWHHNPWTACNVSCG-GGIQPRSPVCVEHSMPCHIL 444
           :  |          | |    | ||| |::|| ||  ||:  :
Db     889 NDQPCSQ---  ---------PVWVTGEWEPCSKSCGKTHMQVRSVKCVQPLNNTTR 936

Qy     445 QVEEWKCMYAPEPEVMQTCNLEPCP KWIAMEWSQCTVSCGRGLKYRVVLCINHRGCEHVC 504
           |   | | :::    || :|| | |||:|||| | : |||      :|  |
Db     937 SVHTEHPCNGARPRSGRKACNRHSCPSWEAGSWGQCSVSCGNGIQERFVLC RTADGSL 994

Qy     505 GCNPQTERNTEKRCVIPEP-EA    PKEF-PVPEAFEDWEPLA   QELLEA    SIA 559
           |  : :      | :  :| :    | :|| |    : || :       :::|    |
Db     995 VCKEE-PPETARTCRLG-PCPRNTGDPSKEGYV   VQWLSSTGPNSFVQEICSKCSCQC 1052

Qy     560 EEPTE----------IPEPWSACSTTCGPGVQVREVKCRV---    --- TTTPTQTE 601
           ::  |           ||    |  :| |   :|  |                | | :
Db     1053 DEGV-PMEAVLPE-LPE-TNETCEA-TGTDENCTLAL-AAT-PNGLPTEMSWTLA 1109
```

```
Qy     602 TELPEEECEGPKLPTERP---CLLEACDESPASRELDIP 637
           | |    | :| | |         | :: | :  ::|
Db     1110 PTLVMEVQPPPCIPLPVPLNTSSTNATGDHPEENAVDVP 1158


RESULT   5
ATS3_HUMAN
ID   ATS3_HUMAN     STANDARD;      PRT;  1201 AA.
AC   Q15072;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   ADAMTS-3 precursor (EC 3.4.24. ) (A disintegrin and metalloproteinase
DE   with thrombospondin motifs 3) (ADAM TS 3) (ADAM TS3) (Fragment).
GN   ADAMTS3 OR KIAA0366.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID 9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE Brain;
RX   MEDLINE 97499841; PubMed 9305841;
RA   Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,
RA   Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. VII.
RT   The complete sequences of 100 new cDNA clones from brain which can
RT   code for large proteins in vitro.";
RL   DNA Res. 4:141-150(1997).
CC   -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR
CC       MATRIX (BY SIMILARITY).
CC   -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
CC       FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC   -!- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY
CC       SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC    !  SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC    !  SIMILARITY: CONTAINS 4 TSP TYPE 1 DOMAINS.
CC    !  CAUTION: HAS SOMETIMES BEEN REFERRED TO ADAMTS4.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AB023451; BAA82821.1;  -.
DR   MEROPS; M12.220;  -.
DR   MIM; 605011;  -.
DR   InterPro; IPR001762; Disintegrin.
DR   InterPro; IPR002870; Peptidase_M12B propep.
DR   InterPro; IPR001590; Reprolysin.
DR   InterPro; IPR000884; TSP1.
DR   InterPro; IPR000140; Zn_Mlpeptdse.
DR   Pfam; PF01562; Pep_M12B_propep; 1.
DR   Pfam; PF01421; Reprolysin; 1.
DR   Pfam; PF00090; tsp_1; 4.
DR   SMART; SM00209; TSP1; 4.
DR   PROSITE; PS50215; ADAM_MEPRO; 1.
DR   PROSITE; PS50092; TSP1; 2.
DR   PROSITE; PS00142; ZINC_PROTEASE; FALSE_NEG.
DR   PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
KW   Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW   Repeat; Extracellular matrix; Heparin-binding.
FT   NON_TER       1      1
FT   SIGNAL       -1      2       POTENTIAL.
FT   PROPEP        2    245       BY SIMILARITY.
FT   CHAIN       246   1201       ADAMTS-3.
FT   METAL       399    399       ZINC (CATALYTIC) (BY SIMILARITY).
```

```
FT   ACT_SITE      307      307       BY SIMILARITY.
FT   METAL         406      406       ZINC (CATALYTIC) (BY SIMILARITY).
FT   METAL         410      410       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DOMAIN        466      546       DISINTEGRIN-LIKE.
FT   DOMAIN        547      603       TSP TYPE 1 1.
FT   DOMAIN        604      708       CYS RICH.
FT   DOMAIN        709      846       EGF-LIKE.
FT   DOMAIN        841      898       TSP TYPE 1 2.
FT   DOMAIN        899      961       TSP TYPE 1 3.
FT   DOMAIN        962     1014       TSP TYPE 1 4.
FT   DOMAIN        242      245       POLY ARG.
FT   CARBOHYD      79        79       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      115      115       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      238      238       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      341      341       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      471      471       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      810      810       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      938      938       N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   1204 AA;  [illegible] MW;  [illegible] CRC64;

  Query Match             [illegible]%;  Score [illegible];  DB 1;  Length [illegible];
  Best Local Similarity   [illegible]%;  Pred. No. [illegible];
  Matches [illegible];  Conservative [illegible];  Mismatches [illegible];  Indels [illegible];  Gaps [illegible];

Qy      96 FEFALSPQGSFLEDI  IGPLFLIYRYIDQISRNITSLEIKDINKYNASLASYSIKTN 155
           |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db     509 FYFCKLEKGDNFKIR  WYKGAV MWRNAGLIKCCVSWGNENCDQERIGSEGMFPD 568

Qy     156 SSASYSFRGVK  SKNFYKLVSNECPPIAEGFNATLPMVNLSTPVNPSLEFQYLN 211
           |  |  |||      |  |  |  |  |  |  |  |||||||  |  |  |  |
Db     569 EVFERFIMQVCNNIPMFINGSLAKSVNFEYGLNTELCLKHFEDERKAGAPVGRNCIEFLN 628

Qy     212 HYYEWLP KYNSFAAIATALKPHAQGNLVVLLAFKVLGRFRCN IEGLDMFTSGIGQAVI 271
           |  |||  |  ||  |  |  |  |  |  |  |||  |  |  |  |  |  |  |  |
Db     629 IKRDWEIYEHPQFKKRTRLYLGKEFIGDVAYMKQLAVRGTTEVIKPYEYGLVKVI 684

Qy     272 PDPGKSNAKFLNFTVTATGVEIVESVEGPKCHVSIEKIGD NVIAVELGSKVKI  327
           |  |||  ||  |||  |  |  |  |  |  |  |  |  |  |  |  |  |
Db     685 PDKFTGSNEVEIPSFSVPSGNEFAVSLETEPTDGIQLNLEPGALSMTLFFVNGVLIERT 744

Qy     328 VKGPAHEFETDSKIDGKSHDEGDHSENSTG      VEVVIENTVSPQFSETAFSTQSDS 383
           |  |  |  |  |  |  |  |  |  |  |      |  |  |  |  |  |  |  |
Db     745 EASPHILAIKNQ    AIGHYTINGKSCKHAKGRTFIDLGVEWDYNIDLPFESENITGIK 798

Qy     384 ADFI            PKTSVIAAKLSVVLYPY  PDLWRLYTFSVTHSVASKTNEIV 433
           |               |  |  |  |||  |  |  |  |  |  |  |  |  |  |
Db     799 DPVIVLIDQENHIRCSILYKYSIHPDSVPLINSNNVIQPELLPEPTWALKSWSLVSKP 858

Qy     434 GGYQLNFAVVSTLPKPVVSTHYTHYSTENVKPKPSTEPSDNSTSGLG SKEIMETIHE 489
           |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db     859 GGPQYIKYGGKRKRCSNKMVHPPE  ANKKPKPIKRMCNTQETHR          904

Qy     490 QMELPRWEHNPWIAVSVSTLAVPLFVVFESMEGETLQVEIRWKCMYAFKPKVM 543
           |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db     905 PLWVAEEWEHCTKTCNSGYQLRIVPCLQPLLNGNCVHGSKYM  PPSSTMPCN 958

Qy     544 LPDCP KWIAMPWSQCTVICGMGLRYRVVLCINHRGRHVGGCNPQLKLHIKEECVTPTPC 603
           |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db     959 RVPCPAQWKTGPWSECSVTCGSGIEVRQVLC  RASGHECGEKPE  SVRACVLP PC 1017

RESULT   6
AIS1_RAT
ID   AIS1_RAT        STANDARD;      PRT;   967 AA.
AC   Q9WUQ1; Q9ERI1;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   ADAMTS-1 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE   with thrombospondin motifs 1) (ADAM-TS 1) (ADAM-TS1).
GN   ADAMTS1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY; TISSUE=Brain;
RA   Liu X., Hu Y., Yin J., Johnstone E.M., Stephenson D.T., Clemens J.A.,
RA   Little S.P.;
RT   "Induction of a disintegrin and metalloprotease with the
RT   thrombospondin type 1 motif (ADAMTS).";
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 18-967 FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Liver;
RX   MEDLINE=20304099; PubMed-10847486;
RA   Dilaimitis I., Luethi M., Hoesli M., Reichen J.;
RT   "Cloning of the rat ADAMTS-1 gene and its down-regulation in
RT   endothelial cells in cirrhotic rats.";
RL   Liver 20:165-172(2000).
CC   -!- FUNCTION: CLEAVES AGGRECAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE
CC       INVOLVED IN ITS TURNOVER. HAS ANTIANGIOGENIC INHIBITOR ACTIVITY (BY
CC       SIMILARITY). ACTIVE MOIETY IS 65 KDA, WHICH MAY BE ASSOCIATED WITH
CC       VARIOUS INFLAMMATORY PROCESSES AS WELL AS DEVELOPMENT OF CANCER
CC       CACHEXIA. MAY PLAY A CRITICAL ROLE IN FOLLICULAR RUPTURE (BY
CC       SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: CLEAVAGE AGGRECAN AT THE [illegible] GLU-[illegible] BOND
CC       SITE, WITHIN THE CHONDROITIN SULFATE ATTACHMENT DOMAIN.
CC   -!- COFACTOR: ZINC (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: SECRETED, ASSOCIATED WITH THE EXTRACELLULAR
CC       MATRIX (BY SIMILARITY).
CC   -!- INDUCTION: DOWN REGULATED IN ENDOTHELIAL CELLS DERIVED IN
CC       CIRRHOTIC LIVER.
CC   -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE 1 2 MOTIF ARE THE PLANT
CC       FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC   -!- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY
CC       SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC   -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 5 TSP TYPE 1 DOMAINS.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; AF149116; AAL34011.1; -.
DR   EMBL; AF034446; AAC29824.1; -.
DR   InterPro; IPR001762; Disintegrin.
DR   InterPro; IPR002870; Peptidase_M12B_propep.
DR   InterPro; IPR001590; Peptidase_M12B.
DR   InterPro; IPR000884; TSP1.
DR   InterPro; IPR000152; Zn_Mipeptidase.
DR   Pfam; PF01562; Peptidase_M12B_propep; 1.
DR   Pfam; PF01421; Reprolysin; 1.
DR   Pfam; PF00090; tsp_1; 5.
DR   SMART; SM00209; TSP1; 5.
DR   PROSITE; PS50215; ADAM_MEPRO; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
DR   PROSITE; PS00209; TSP1; 2.
DR   PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
KW   Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW   Repeat; Extracellular matrix; Heparin-binding.
FT   SIGNAL        1     54       POTENTIAL.
FT   PROPEP       55    252       BY SIMILARITY.
FT   CHAIN       253    967       ADAMTS-1.
FT   SITE        295    296       CYSTEINE SWITCH (POTENTIAL).
FT   METAL       401    401       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    402    402       BY SIMILARITY.
FT   METAL       405    405       ZINC (CATALYTIC) (BY SIMILARITY).
FT   METAL       411    411       ZINC (CATALYTIC) (BY SIMILARITY).
```

```
FT   DOMAIN       476     558       DISINTEGRIN LIKE.
FT   DOMAIN       559     615       TSP TYPE 1 1.
FT   DOMAIN       616     724       CYS RICH.
FT   DOMAIN       725     867       SPACER.
FT   DOMAIN       868     907       TSP TYPE 1 2.
FT   DOMAIN       908     967       TSP TYPE 1 3.
FT   DOMAIN       194     198       POLY ARG.
FT   CARBOHYD     547     547       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     724     724       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     764     764       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     782     782       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     945     945       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT      21      21       I -> V (IN REF. 2).
FT   CONFLICT      26      41       KPRRPG -> RRRRPP (IN REF. 2).
FT   CONFLICT      49      49       V -> A (IN REF. 2).
FT   CONFLICT      72      72       R -> P (IN REF. 2).
FT   CONFLICT      79      79       T -> TP (IN REF. 2).
FT   CONFLICT     249     249       R -> G (IN REF. 2).
FT   CONFLICT     262     265       TMEV -> NELK (IN REF. 2).
FT   CONFLICT     607     607       S -> P (IN REF. 2).
FT   CONFLICT     936     936       L -> V (IN REF. 2).
FT   CONFLICT     962     962       I -> T (IN REF. 2).
SQ   SEQUENCE   967 AA;  105705 MW;  F93C864F6DCDB4CF CRC64;


  Query Match             6.3%;  Score 584.5;  DB 1;  Length 967;
  Best Local Similarity   27.7%;  Pred. No. 2e-29;
  Matches  132;  Conservative   75;  Mismatches  176;  Indels   93;  Gaps   16;

Qy      76 GNWDAWGDWSECSRTCGGGASYSLRKCLE------GRNCSGQNIPYKICSNHDCPPDGAE 128
           |:|  || | ||||||||||  |:|| |        |: ||| :|||:|:  |||  : :
Db     560 GSWGPWTPWTLCSRTCGVGVTTMEPVCNNETVEPSGDKTTLCEPVEYEAKNITLQCPTNNSP 619

Qy     129 PERAQQCSAYNCVQYQH    YYEWLPEYN  DPAAPCALKCHAQCQNLVVELAPKVLD 182
             || ||| |:|:          || |:|    |   | | | |:|       | |||:|
Db     620 TFSIEQCTAHNISFAFNNTSKNFTVEWTFYAGVKYLLFKEVPVSGLTSESVKTGSLLVQFKVVL 679

Qy     183 GTRCNTLSLDMCTSVCQAVDGHQLGVNAKEQNGVVCAWDGPEAVRQLTKFHVLPLF 240
           :| |: ||  :|: | |  :||| : |  | | |||| |:|||: : |   : |
Db     680 GTPCSPDSTSVCVQQCVEAVCDPIDSKKKFDSKCVCNPGYSTCKKIGH  TVTCEPP 736

Qy     241 REENVIAVELCTRSVRCTVPDAHETTPWTVQFFTHCP    NETCVEVVPNTTVET 298
           ::: :| |: :: : :|          :||:  |:     : | :::      |
Db     737 GTHDIVTTFAGATNIEVKEHNP    :  :  RGSRNRCSTLAIRAALLTYTLNG   NE 792

Qy     299 QRQSFRQTTPKIPGPLMADFIPKTPYTAAKDSVVQPTPYYQPLSHQWPQTDPTPTTV-IQGS 357
           :  |    | ::       ||: :  :: :   : |:          | |:
Db     784 ILSILEQDLIYETIVL -    PYSGSSAAILEKECLGLSIEE        F..:|V.HV 833

Qy     358 GYGLNSAEVCKSLPPVVNGHYCHETTFLRVEEPEPLTLMHPLPGLSDSCMEAMRLLQLS 417
           |: |                        :|||:|             |:  |::
Db     829 GHAL----------------  :  :    RPKIKYTYTPMK       EKKIEPINAIP 874

Qy     418 PLPPWPHNPWTAQSVPQCCQLQRPPVQVPPSMHCPIQQVPEEWETMYAPEPEVMQTQNLI 477
           |    | :|  :|: | |||   | : :|         :|    ||   : |
Db     875 IPSEWVIEEWGECSKTCGSGWQRPVVECRDINGH-  -- PASQCAKPVKPASTRPCADL 927

Qy     478 QCEEWSAWSEWS-TCVTCGTGVETRVVLCINSDCLLVK  QKSCDLE LPEPETRVI 536
           ||:|   :|| |: |||:| : | : |:: | :    :|  |  |  : |::
Db     928 PPPPWQYTWTPRGPCSRSCGGGVQFRYRLCVSHEDCVLDE TLIEPSFCNT---EKP 984


RESULT   7
ATS1_HUMAN
ID   ATS1_HUMAN     STANDARD;      PRT;   967 AA.
AC   Q9UHI8; Q9UP90; Q9UHB4; Q9P2K8; Q9NSJ9;
DT   30-MAY-2000 (Rel. 39, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   ADAMTS-1 precursor (EC 3.4.24. ) (A disintegrin and metalloproteinase
DE   with thrombospondin motifs 1) (ADAM TS 1) (ADAM TS1) (METH 1).
GN   ADAMTS1 OR METH1 OR KIAA1346.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Casas C., Pritchard M.A., Estivill X., Arbones M.L.;
RC   "Cloning, characterization and mapping on human chromosome 21 of the
RC   orthologue of murine Adamts-1.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A., AND PARTIAL.
RC   TISSUE=Heart;
RX   MEDLINE=99367466; PubMed=10438512;
RA   Vazquez F., Hastings G., Ortega M.A., Lane T.F., Oikemus S.,
RA   Lombardo M., Iruela-Arispe M.L.;
RT   "METH-1, a human ortholog of ADAMTS-1, and METH-2 are members of a new
RT   family of proteins with angio-inhibitory activity.";
RL   J. Biol. Chem. 274:23349-23357(1999).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Endothelial cells;
RX   MEDLINE=20247184; PubMed=10785405;
RA   Glienke J., Schmitt A.O., Pilarsky C., Hinzmann B., Weiss B.,
RA   Rosenthal A., Thierauch K.H.;
RT   "Differential gene expression by endothelial cells in distinct
RT   angiogenic states.";
RL   Eur. J. Biochem. 267:2820-2830(2000).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=20181126; PubMed=10718198;
RA   Nagase T., Kikuno R., Ishikawa K.-I., Hirosawa M., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes.XVI.
RT   The complete sequences of 150 new cDNA clones from brain which code
RT   for large proteins in vitro.";
RL   DNA Res. 7:65-73(2000).
RN   [5]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20289799; PubMed=10830953;
RA   Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
RA   Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,
RA   Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,
RA   Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
RA   Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
RA   Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
RA   Shintani A., Sasaki T., Nagamine K., Mitreva M., Wilson R.K.,
RA   Minoshima S., Shimizu N., Nordsiek G., Hornischer K., Brandt P.,
RA   Scharfe M., Schon O., Desario A., Reichelt J., Kauer G., Blocker H.,
RA   Ramser J., Beck A., Klages S., Hennig S., Riesselmann L., Dagand E.,
RA   Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
RA   Lehrach H., Reinhardt R., Yaspo M.-L.;
RT   "The DNA sequence of human chromosome 21.";
RL   Nature 405:311-319(2000).
RN   [6]
RP   SEQUENCE OF 418-967 FROM N.A.
RC   TISSUE=Melanoma;
RA   Blum H., Bauersachs S., Mewes H.-W., Weil B., Wiemann S.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: CLEAVES AGGRECAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE
CC       ANGIOGENIC IN ITS OWN RIGHT (BY SIMILARITY). AN ANTIONCOGENIC
CC       ACTIVITY, ACTIVE METALLOPROTEASE, WHICH MAY BE ASSOCIATED WITH
CC       VARIOUS INFLAMMATORY PROCESSES AS WELL AS DEVELOPMENT OF CANCER
CC       CACHEXIA, MAY PLAY A CRITICAL ROLE IN FOLLICULAR RUPTURE.
CC   -!- CATALYTIC ACTIVITY: CLEAVES AGGRECAN AT THE 1938-GLU-|-LEU-1939
CC       SITE, WITHIN THE CHONDROITIN SULFATE ATTACHMENT DOMAIN.
CC    !  COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: SECRETED, ASSOCIATED WITH THE EXTRACELLULAR
CC       MATRIX (BY SIMILARITY).
CC    !  DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE 1 DOMAINS ARE IMPORTANT
CC       FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC    !  PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY
CC       SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
```

```
CC   !  SIMILARITY: CONTAINS 1 DISP TYPE I DOMAINS.
CC   !  SIMILARITY: CONTAINS 1 TSP TYPE I DOMAINS.
CC
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.[illegible]/sg_[illegible]
CC   or send an email to license@isb-sib.ch).
CC
DR   EMBL; AF170084; AAF15617.1; -.
DR   EMBL; AF060152; AAD38060.1; ALT_INIT.
DR   EMBL; AF207664; AAF23772.1; -.
DR   EMBL; AB037677; BAA92584.1; ALT_INIT.
DR   EMBL; AF001567; BAA95562.1; -.
DR   EMBL; AL162080; CAB82415.1; -.
DR   MIM; 606174; -.
DR   InterPro; IPR001762; Disintegrin.
DR   InterPro; IPR002870; Pep_M12B_propep.
DR   InterPro; IPR001590; Reprolysin.
DR   InterPro; IPR000884; TSP1.
DR   InterPro; IPR000152; Zn_Mlprtase.
DR   Pfam; PF01562; Pep_M12B_propep; 1.
DR   Pfam; PF01421; Reprolysin; 1.
DR   Pfam; PF00090; TSP_1; 1.
DR   SMART; SM00050; DISIN; 1.
DR   SMART; SM00209; TSP1; 1.
DR   PROSITE; PS50215; ADAM_MEPRO; 1.
DR   PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
DR   PROSITE; PS50092; TSP1; 2.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW   Repeat; Extracellular matrix; Heparin binding.
FT   SIGNAL        1     40       POTENTIAL.
FT   PROPEP       50    252       BY SIMILARITY.
FT   CHAIN       253    962       ADAMTS 1.
FT   DISULFID    198    198       CYSTEINE SWITCH (POTENTIAL).
FT   METAL       401    401       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    402    402       BY SIMILARITY.
FT   METAL       405    405       ZINC (CATALYTIC) (BY SIMILARITY).
FT   METAL       411    411       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DOMAIN      476    559       DISINTEGRIN-LIKE.
FT   DOMAIN      560    616       TSP TYPE I 1.
FT   DOMAIN      617    674       CYS-RICH.
FT   DOMAIN      675    849       SPACER.
FT   DOMAIN      850    908       TSP TYPE I 2.
FT   DOMAIN      909    967       TSP TYPE I 3.
FT   DOMAIN      943    946       POLY-LYS.
FT   CARBOHYD    547    547       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    720    720       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    764    764       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CONFLICT    227    227       S -> A (IN REF. 4 AND 5).
FT   CONFLICT    408    408       L -> H (IN REF. 1).
FT   CONFLICT    561    561       S -> N (IN REF. 1).
SQ   SEQUENCE   967 AA;  105303 MW;  C1DC0893247411FD1 CRC64;


  Query Match             5.1%;  Score 571.5;  DB 1;  Length 967;
  Best Local Similarity   28.2%;  Pred. No. 1.6e-28;
  Matches  136;  Conservative  [illegible]  Mismatches  364;  Indels  [illegible];  Gaps  [illegible];

Qy      76 GNWDAWGDWSDCSRTCGGGASYSSRDCLE      GRNCEGQNIRYKICSNHRCPE IAP 228
           |:|  || | ||||||:|| |:|:| |        |: ||| :|||:|: |  :  .
Db     560 GSWGMWSPWGDCSRDCGGVQYTMRSCDNPVPKNGGKYCEGKRVRYRSCNIPDCPDNSK 619

Qy     229 DFRAQQCGAYNIDVDY    QGHYYEWLRPYN  DRAAPVALEVHAVDGLNEVVECAREVFL 282
           || ||| |:|:          :|:||: |  | | :|  :  |  |:|:|:|:|:|:|:|:
Db     620 DFREEQCLAHNEFSKASFGSSVEWIPKYAGVSPKDPKLCQARDQGYEFVFCLPDVVK 676

Qy     283 GTRCNIHSLDMCISGICQVAVCDGLGNAKEDNCGVCAGDNSCCKLVRGQGKSHVSFEK 342
           || | || :|:  |    ||  : |  | | | | |:|||:|:  :          :
Db     677 SIPCSDRGVSTKVVQGVKAQLKEHSKPERVCVCSNGSTPKKGS   SVTSAS  720
```

```
Qy     343 RCLNVTAVFGGNKVKLVRGIAHLEISKLIQGSKGEHSF  SNSPGVFVVLN  402
           :|| :| |:::          :: :  :||:  ||    | |:::     :
Db     737 GYHDLIIIPIGAINIE          VKQRNQRGSRNNGSFLAIKAADGTYILNGDYTLS 786

Qy     403 TVEPQRGSERQTPKTIPGPLMALSTPK  IRYTIAAKDSVVQETTYQPTSHQWRQTDPPSC 462
           |:|                  | :||   ||: :  :: :  :  :           |
Db     787 TLE                QEIMYKGVVLRYSGSSAALERIRSFSPLKE        PL 821

Qy     463 CA  VAT SCG  R  S  S  R RVVC  G  GVQFRNVRTGDTGRLKEIM 511
           ::       |    : :::      | ::          | |: : :  |:
Db     822 CLQV     LIVGNALRPKIK       YIYFV         KKKKESENAIPI          854

Qy     512 PYSHPAPQFGQLWLHNEWIATSVSPGSVFVDFLVTSGPMHGLNLSAWHDKTMYAPKFKVM 571
                   |    |  || ||  | :|:  :    :|:     :|    :|    ||
Db     855         FSAWVIEEWGECSKSCGLGWQMREVRC  RDINSQFAS    PGAKECVKPASE 901

Qy     572 QFCNEETKPEWIAMEWSQCTVTCGGGLRYRVVKCINHHGEEVS  GCNPQLK PITKEFC 628
           : :   |||  |||  | :|:|: | | :::  | |  :  :  :  :  |  | : :
Db     902 RPCADHPCQWQLGEWSSCSKSCGKGYKKRSLKCIDHGGEVSHESCSGPLKKCKHEHDC 961

Qy     629 VL 630
           .
Db     962 IM 963


RESULT    8
AIS1_MOUSE
ID   AIS1_MOUSE     STANDARD;      PRT:   968 AA.
AC   P97857; Q54768;
DT   30-MAY-2000 (Rel. 39, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   ADAMTS-1 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE   with thrombospondin motifs 1) (ADAM-TS 1) (ADAM-TS1).
GN   ADAMTS1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129/SVJ;
RX   MEDLINE=98110584; PubMed=9441751;
RA   Kuno K., Iizasa H., Ohno S., Matsushima K.;
RT   "The exon/intron organization and chromosomal mapping of the mouse
RT   ADAMTS-1 gene encoding an ADAM family protein with TSP motifs.";
RL   Genomics 46:466-471(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97159761; PubMed=8995253;
RA   Kuno K., Kanada N., Nakashima E., Fujiki F., Ichimura F.,
RA   Matsushima K.;
RT   "Molecular cloning of a gene encoding a new type of metalloproteinase-
RT   disintegrin family protein with thrombospondin motifs as an
RT   inflammation associated gene.";
RL   J. Biol. Chem. 272:556-562(1997).
RN   [3]
RP   CHARACTERIZATION, AND METAGENESIS OF GUT.
RX   MEDLINE=97036657; PubMed=8973500;
RA   Kuno K., Terashima Y., Matsushima K.;
RT   "ADAMTS-1 is an active metalloproteinase associated with the
RT   extracellular matrix.";
RL   J. Biol. Chem. 274:19963-19966(1999).
RN   [4]
RP   FUNCTION.
RX   MEDLINE=20389568; PubMed=10954706;
RA   Kuno K., Okada Y., Kawashima H., Nakamura H., Miyasaka M.,
RA   Ohno H., Matsushima K.;
RT   "ADAMTS-1 cleaves a cartilage proteoglycan, aggrecan.";
RL   FEBS Lett. 478:241-245(2000).
RN   [5]
RP   FUNCTION, AND INHIBITION.
RX   MEDLINE=20243757; PubMed=10781075;
```

```
RA   Bekker R.I., Russell D.L., Espey L.L., Lyden J.P., O'Malley B.W.,
RA   Richards J.S.;
RT   "Progesterone regulated genes in the ovulation process: ADAMTS 1 and
RT   cathepsin L proteases.";
RL   Proc. Natl. Acad. Sci. U.S.A. 97:4689-4694(2000).
CC   -!- FUNCTION: CLEAVES AGGRECAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE
CC       INVOLVED IN ITS TURNOVER. HAS ANGIOGENIC INHIBITOR ACTIVITY (BY
CC       SIMILARITY). ACTIVE METALLOPROTEASE, WHICH MAY BE ASSOCIATED WITH
CC       VARIOUS INFLAMMATORY PROCESSES AS WELL AS DEVELOPMENT OF CANCER
CC       CACHEXIA. MAY PLAY A CRITICAL ROLE IN FOLLICULAR RUPTURE (BY
CC       SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: CLEAVES AGGRECAN AT THE 1691-GLU-|-ILE-1692
CC       SITE, WITHIN THE CHONDROITIN SULFATE ATTACHMENT DOMAIN.
CC   -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: SECRETED, ASSOCIATED WITH THE EXTRACELLULAR
CC       MATRIX.
CC   -!- INDUCTION: INDUCED IN VITRO IN COLON ADENOCARCINOMA CELLS BY
CC       INTERLEUKIN 1, TNF IN VIVO IN KIDNEY AND HEART BY
CC       LIPOPOLYSACCHARIDE. ALSO INDUCED BY LH STIMULATION IN GRANULOSA
CC       CELLS OF PREOVULATORY FOLLICLES
CC   -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
CC       FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC   -!- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC   -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 4 TSP TYPE-1 DOMAINS.
CC   -!- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
CC       FRAMESHIFT IN POSITION 7.
CC   ---------------------------------------------------------- ---- ---- ----------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------- ---- ---- ----------
DR   EMBL; AF201735; AAM34562.1; ALT_INIT.
DR   EMBL; D67076; BAA11088.1; ALT_FRAME.
DR   MEROPS; M12.222; -.
DR   MGD; MGI:109249; Adamts1.
DR   InterPro; IPR001762; Disintegrin.
DR   InterPro; IPR002870; Pep_M12B_propep.
DR   InterPro; IPR001590; Reprolysin.
DR   InterPro; IPR000884; TSP1.
DR   InterPro; IPR000140; Zn_Mlpeptidse.
DR   Pfam; PF01562; Pep_M12B_propep; 1.
DR   Pfam; PF01421; Reprolysin; 1.
DR   Pfam; PF00090; tsp_1; 3.
DR   SMART; SM00209; TSP1; 3.
DR   PROSITE; PS50215; ADAM_MEPRO; 1.
DR   PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
DR   PROSITE; PS50092; TSP1; 3.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW   Repeat; Extracellular matrix; Heparin binding.
FT   SIGNAL       1     48       POTENTIAL.
FT   PROPEP      49    254
FT   CHAIN      255    968       ADAMTS 1.
FT   SITE       256    256       CYSTEINE SWITCH (POTENTIAL).
FT   METAL      402    402       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE   403    403
FT   METAL      406    406       ZINC (CATALYTIC) (BY SIMILARITY).
FT   METAL      412    412       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DOMAIN     477    559       DISINTEGRIN-LIKE.
FT   DOMAIN     560    617       TSP TYPE-1 1.
FT   DOMAIN     618    725       CYS-RICH.
FT   DOMAIN     726    850       SPACER.
FT   DOMAIN     851    909       TSP TYPE-1 2.
FT   DOMAIN     910    968       TSP TYPE-1 3.
FT   DOMAIN     195    199       POLY-ARG.
FT   CARBOHYD   548    548       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   721    721       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   765    765       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   783    783       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   946    946       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   MUTAGEN    403    403       E->Q: LOSS OF ACTIVITY.
FT   CONFLICT   335    335       N -- S (IN REF. 2).
FT   CONFLICT   425    425       T -- S (IN REF. 2).
SQ   SEQUENCE   968 AA;  105841 MW;  42EBDA55499FB6C1 CRC64;


  Query Match             6.0%;  Score 564.5;  DB 1;  Length 968;
  Best Local Similarity   27.7%;  Pred. No. 3.8e-28;
  Matches  137;  Conservative   72,  Mismatches  173;  Indels  103;  Gaps   18;

Qy      76 CNWDAWCPWSDCSRTCGGGASYSLRFCLT      GRNPPDSQNIRYKTCNNHPCFP DAP 128
           |:|  || | ||||||||| |::| |        |: ||| :||:::|: ||| : :
Db     561 GSWGPWGPWGPCSRTCGGGVQYTMRECDNPVPKNGGKYCPGKRVRYRSCNIEDCPDNNGK 620

Qy     129 LPAQQTSAYNPVQYQM ---YYPWIPRYN- LPAAPTALPCHAQSQNLVVELAPKVLL 182
           || :|| |:|:        || |:|   |   | | | |:|     | ||||:|
Db     621 TPPEEQTPAHNFSPASPCNPDTVPWIPKYAGVSPKDRCKTICPAPCIGYFEVIGPKVVD 680

Qy     183 GTRCNTDSLDMCISGICQAVGCDRQLGSNAEDNCGVCAGDGSICRLVRGQSKSHVSPEK 242
           || | || :|| | |   |||| : |  | | |||| :|||:: : |      |
Db     681 SIPCPIPVNVVSGIDVPEAGTDEIIVTPPEIPPPVVPNPSPCPKMSG   IVTSTRP 737

Qy     243 FEPNVIAVPLGSPVPFIVPGPAHLPTISPTLQGSKQEHST    NSPGVPVVENTTVPP 298
           ::: :| |: ::          :: :  :||:   ||     : | :::      |
Db     738 GTRDIVTITAGAINIE          VKRRNQLRGSRNNGSFLAIKAADATTYILNG   NP 784

Qy     299 QPGSPMQTPPQTPTGTV    MAPPTPPPPPAAPSVVGPSPVQPISPQWRQTDFPPCIV-TGGG 357
           :  |    | ::       ||: : :: :  :|:           | |:
Db     785 TIRTIFQGTPPQTVI    RTSGPSAAIPRIRPSPSIFP--------PLTIQVLMV 829

Qy     358 CNGAPCVLIPLPVVLNGSCHYPLSAVPLTPLPPLSMQPCLPSPTPLLM  PYLH 415
           |: |                        :||:|          | |  :
Db     830 GRAI----  --    ---  ----------RPNP----------EIYEMKKIES 849

Qy     416 FQPLP   PWEHRPWTAPSVSPTPTQPRSIVPVLESMELPIIL_VIPWPCMYAIPLPVM 472
           |  :|    |    | ||:|| | ||   | : : |          :|    ||   :
Db     850 PNAIPTPPSEWVIPEWGPCSKTCGSGWQRRVVQCRDINGH       PASETAKEVKPASTR 903

Qy     473 TCNLPSPAPIDNWQLVAGWSPCSRSCGMGVVTRNRMCLHVA  SGDQLAK SRPEEKPV 529
           |    || |  :|| | : | : :|:|  :   :    :|:| | |  | |  : |
Db     904 PCADLPCPNWQVGNWSPCSKSCGKGYRTRSRMCVDIN GSVLDNS CPLEKPEHYIDPCT 963

Qy     530 L 530
           :
Db     964 L 964


RESULT   9
ATS2_HUMAN
ID   ATS2_HUMAN     STANDARD;      PRT;  1211 AA.
AC   O95450;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2002 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   ADAMTS-2 precursor (EC 3.4.24.14) (A disintegrin and
DE   metalloproteinase with thrombospondin motifs 2) (ADAM-TS 2) (ADAM-TS2)
DE   (Procollagen I/II amino propeptide processing enzyme) (Procollagen I
DE   N proteinase) (PC I NP) (Procollagen N endopeptidase) (pNPI)
DE   (Procollagen I/II amino-propeptide processing enzyme).
GN   ADAMTS2 OR PCINP OR PCPNI.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS 1PNPI AND 2PNPI).
RC   TISSUE=Skin;
RX   MEDLINE=99347935; PubMed=10417273;
RA   Colige A., Sieron A.L., Li S. W., Schwarze U., Petty E.,
RA   Wertelecki W., Wilcox W., Krakow D., Cohn D.H., Reardon W.,
```

```
RA   Byers P.H., Lapiere C.M., Prockop D.J., Nusgens B.V.;
RT   "Human Ehlers-Danlos syndrome type VII C and bovine dermatosparaxis
RT   are caused by mutations in the procollagen I N-proteinase gene.";
RL   Am. J. Hum. Genet. 65:308-317(1999).
CC   -!- FUNCTION: CLEAVES THE PROPEPTIDES OF TYPE I AND II COLLAGEN PRIOR
CC       TO FIBRIL ASSEMBLY. DOES NOT ACT ON TYPE III COLLAGEN. MAY ALSO
CC       PLAY A ROLE IN DEVELOPMENT THAT IS INDEPENDANT OF ITS ROLE IN
CC       COLLAGEN BIOSYNTHESIS.
CC   -!- CATALYTIC ACTIVITY: Cleaves the N-propeptide of collagen chain
CC       alpha 1(I) at Pro-|-Gln and of alpha 1(II) and alpha 2(I) chains
CC       at Ala-|-Gln.
CC   -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC   -!- SUBUNIT: MAY BELONG TO A MULTIMERIC COMPLEX. BINDS SPECIFICALLY TO
CC       COLLAGEN TYPE XIV (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR
CC       MATRIX (BY SIMILARITY).
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; LPNPI (SHOWN HERE) AND SPNPI.
CC       ARE PRODUCED BY ALTERNATIVE SPLICING. SPNPI HAS NO SIGNIFICANT N
CC       PROCOLLAGEN PEPTIDASE ACTIVITY.
CC   -!- TISSUE SPECIFICITY: EXPRESSED AT HIGHT LEVEL IN SKIN, BONE, TENDON
CC       AND AORTA AND AT LOW LEVELS IN THYMUS AND BRAIN.
CC   -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
CC       FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC   -!- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY
CC       SIMILARITY).
CC   -!- DISEASE: DEFECTS IN ADAMTS2 ARE THE CAUSE OF EHLERS-DANLOS
CC       SYNDROME TYPE VIIC (EDS VIIC), A RECESSIVELY INHERITED DISORDER
CC       CHARACTERIZED CLINICALLY BY SEVERE SKIN FRAGILITY AND
CC       BIOCHEMICALLY BY THE PRESENCE IN SKIN OF PROCOLLAGEN INCOMPLETELY
CC       PROCESSED AT THE AMINO TERMINUS.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC   -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 4 TSP TYPE-1 DOMAINS.
CC   -!- CAUTION: HAS SOMETIMES BEEN REFERRED TO ADAMTS3.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for  commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AJ003125; CAA05880.1; -.
DR   MEROPS; M12.301; -.
DR   MIM; 604539; -.
DR   MIM; 225410; -.
DR   InterPro; IPR001762; Disintegrin.
DR   InterPro; IPR002870; Pep_M12B_propep.
DR   InterPro; IPR001590; Reprolysin.
DR   InterPro; IPR000884; TSP1.
DR   InterPro; IPR000130; Zn_MIpeptdse.
DR   Pfam; PF01562; Pep_M12B_propep; 1.
DR   Pfam; PF01421; Reprolysin; 1.
DR   Pfam; PF00090; tsp_1; 1.
DR   SMART; SM00209; TSP1; 4.
DR   PROSITE; PS50215; ADAM_MEPRO; 1.
DR   PROSITE; PS50092; TSP1; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; FALSE_NEG.
DR   PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
KW   Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW   Repeat; Collagen degradation; Extracellular matrix; Heparin-binding;
KW   Alternative splicing.
FT   SIGNAL        1     29       POTENTIAL.
FT   PROPEP       30    253       BY SIMILARITY.
FT   CHAIN       254   1211       ADAMTS 2.
FT   METAL       408    408       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    409    409       BY SIMILARITY.
FT   METAL       412    412       ZINC (CATALYTIC) (BY SIMILARITY).
FT   METAL       418    418       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DOMAIN      480    560       DISINTEGRIN-LIKE.
FT   DOMAIN      561    617       TSP TYPE-1 1.
FT   DOMAIN      618    722       CYS-RICH.
FT   SITE        691    693       CELL ATTACHMENT SITE (POTENTIAL).
FT   DOMAIN      723    851       SPACER.
FT   DOMAIN      852    911       TSP TYPE-1 2.
FT   DOMAIN      912    974       TSP TYPE-1 3.
FT   DOMAIN      975   1030       TSP TYPE-1 4.
FT   DOMAIN       30     43       POLY-ALA.
FT   DOMAIN      185    188       POLY-GLU.
FT   CARBOHYD    112    112       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    251    251       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    949    949       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    993    993       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1031   1031       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1098   1098       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1145   1145       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1150   1150       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC    544    566       HCLKGHCIWLTPLTKRDGSWGA-->ERPGAVAHACYPS
FT                                TLGGQGRWIA (IN ISOFORM SPNPI).
FT   VARSPLIC    567   1211       MISSING (IN ISOFORM SPNPI).
SQ   SEQUENCE   1211 AA;  134722 MW;  B5CEEF2[illegible]23CAD2F [illegible];

  Query Match             6.0%;  Score 564;  DB 1;  Length 1211;
  Best Local Similarity   21.2%;  Pred. No. 6.5e-28;
  Matches  178;  Conservative   88;  Mismatches  233;  Indels  342;  Gaps   26;

Qy         74 K[illegible]NW[illegible]AW[illegible]W[illegible]A[illegible]          [illegible]N[illegible]YK[illegible]NH[illegible]A 127
              :::| ||  :  ||||| |  :  |:|               || | |    :: ||  |||
Db        560 KDGSWGAWSPFGS[illegible]SR[illegible]VKFRTRQ[illegible]DNPHPANG[illegible]SGLAYDFQLCSRQ[illegible]PDSL 619

Qy        128 EDERAQQCSAYN          DVQYQGHYYEWLP[illegible]KYNDPAAPCALKCHAQGQNLVVELAPKV 180
               ||| :||  ::          | |        : |||  : |      | | | ::     || :    |
Db        620 ADFREEQCRQWDLYFEHGDAQ     HHWLPHEH[illegible]DAKE[illegible]G[illegible]VVSMKRMV 674

Qy        181 [illegible]N[illegible]D[illegible]M[illegible]AV[illegible]D[illegible]NAK[illegible]N[illegible]V[illegible]A[illegible]KLVP[illegible]SKSHVS 239
               |||||:    |:   :|:  | | |:  ||||   :||:  :|| |||| || |  | ::|:|        |
Db        675 [illegible]SYKDA[illegible]VR[illegible]KV[illegible]V[illegible]SSKQ[illegible]V[illegible]PNSH[illegible]KVVKGTFTR  S 732

Qy        240 PEKREE-NVIAVPLGSRSVRITVKGPAHLFIESKTLQGSK----GEHSEN-SPGVFVVEN 293
               :|      :  :| |:| : |                :  | | :   |     |:  :  |     |:
Db        733 PKKHGYIKMFEIPAGAPHLLIQEVDATSHHLAVKNLETGKFILNEENDVDASSKTFIAMG 792

Qy        294 TIVEFQRGSERQTFKIPGPLMADFIFKTRYTAAKDSVVQFFFYQPISHQWRQTDFFPCTV 353
                  |::      |:| :   |||
Db        793 VEWEYRDEDGRETLQTMGPL                                         812

Qy        354 [illegible]GGYQLNSAE[illegible]V[illegible]KRVV[illegible]HY[illegible]HYY[illegible]ENVKPKPKLK[illegible]SM[illegible]SS[illegible]GEKEIMPY 413

Db        813                                                              812

Qy        414 [illegible]HFQPLPRWEHNPWTACSVSC[illegible]QRRSFVCVEESMHGEIIQVEEWKCMYAPKPKVMQT 473
                       |     |     :    |          |   :   :   :|  :  |  |::          |::
Db        813           HGTTIVLVIPVG     DTRVSLTYKYMIHEDSLNVDD            NNVLEE 851

Qy        474 CNLF[illegible]PKWIAMEWS[illegible]TV[illegible]YKVVL[illegible]NHRGEHV[illegible]N[illegible]LK[illegible]HIKFE[illegible]VIP[illegible] 533
               :          :  ||:
Db        852 DS         VVYEWA                                            859

Qy        534 [illegible]YKEKEKS[illegible]VLAKLEWLKQAQ[illegible]IA[illegible]E[illegible]WSA[illegible]V[illegible]VEFVKCFV 593
                            ||:                              || ||   || |  |    ||
Db        860               [illegible]KK                              WSP[illegible]SKP[illegible]S[illegible]FIKYC[illegible]R 882

Qy        594 [illegible]EL[illegible]PEEE[illegible]KT[illegible]L[illegible]EACDESPASRELD[illegible]SEITYDWEYAGF 653
                                                            | ||
Db        883                                               RRLD               886

Qy        654 [illegible]TA[illegible]KAIAV[illegible]VN[illegible]MVHPPAMG[illegible]A[illegible]N[illegible]RHV[illegible]W 712
                                       : :    :  :|  |: :||| : |  |    ||
Db        887                   HKMVHRGFCAALSKPKATRRACNPQECSQPVWVTGEW 923

Qy        713 G[illegible]SA[illegible]G  VG[illegible]TRDVY[illegible]      G[illegible]PAP[illegible]KPHAL[illegible]NQ[illegible]PPGWHTEE 767
               ||| |||   |:| | | |:              : |  | :|  : :||::   || |
Db        924 EPCSQTCGRTGMQVRSVRC[illegible]QLLHDNTTRSVHAKHCNGARPESRRACSRELCPGRWRAGP 983
```

```
Qy     768 WQQCSRICQGGGTQNRKVICRQLLLHQSTLNLSHELQGHKASSHKSCAETDCPPHLAVGD 827
           | ||| ||| ||  | |  |     | ||       :||  :  : ::|    || :  : |
Db     984 WSQCSVICGNGTQERPVPCR--TADDSF-----GICQEERPETARTCRLGPCPRN--ISD 1034

Qy     828 WQKCSVECNGGIQYRYRLVCVQRLAAKGRPIPLSEMMEF          DLPGEPLV  RSCQ 876
           || | |   :  :::::||       : ||           :|:  :  :||
Db    1035 PSPEETVVCWLPECTKPSFRCQNGTCVPMDAVCDGIDDCGDQSDEKNCEGVSCN 1089

Qy     877 M 877
           :
Db    1090 L 1090


RESULT  10
ATI0_MOUSE
ID   ATI0_MOUSE     STANDARD;      PRT;   950 AA.
AC   P58459;
DT   01-MAR-2002 (Rel. 41, Created)
DT   01-MAR-2002 (Rel. 41, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   ADAMTS-10 (EC 3.4.24.-) (A disintegrin and metalloproteinase with
DE   thrombospondin motifs 10) (ADAM-TS 10) (ADAM-TS10) (Fragment).
GN   ADAMTS10.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Seldin M.F., Apte S.S.;
RT   "A novel gene of the ADAMTS family predicts ADAMTS-10, a
RT   metalloprotease with unique structural features and expression
RT   pattern.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: SECRETED, ASSOCIATED WITH THE EXTRACELLULAR
CC       MATRIX (BY SIMILARITY).
CC   -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE 1 DOMAINS ARE IMPORTANT
CC       FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX (By
CC       similarity).
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC   -!- SIMILARITY: CONTAINS AT LEAST 4 TSP TYPE 1 DOMAINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AF402012; AAK97226.1; -.
DR   PROSITE; PS50092; TSP1; 1.
KW   Hydrolase; Metalloprotease; Zinc; Glycoprotein; Repeat;
KW   Extracellular matrix.
FT   NON_TER       1       1
FT   DOMAIN       -1      51       CYS-RICH.
FT   DOMAIN       52     174       SPACER.
FT   DOMAIN      171     232       TSP TYPE-1 1.
FT   DOMAIN      234     290       TSP TYPE-1 2.
FT   DOMAIN      294     348       TSP TYPE-1 3.
FT   DOMAIN      353     399       TSP TYPE-1 4.
FT   CARBOHYD     65      65       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    141     141       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    248     248       N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   450 AA;  48861 MW;  2377D0E2CFBFBFCA CRC64;


  Query Match             5.8%;  Score 544.5;  DB 1;  Length 450;
  Best Local Similarity   28.2%;  Pred. No. 2.5e-27;
  Matches  150;  Conservative   62;  Mismatches  196;  Indels  129;  Gaps   22;

Qy     177 APEVILDIAVVMDKSMEMKGATLAVTQALKSVDQLSDNAEFSFVEFNSDAHPLTPSLQPT 236
           |  :  :|  :  :|  :  |  |     : |
Db       1 AAAVVDGTPCRQFDTVDLCVSGGCKHVGCDRVEGSDLDKEDKCKVCGGDGSACETTEGVF-S 59

Qy     237 HVSPEKREENVIAVPLGSRSVRITVRGPAHLFI----------------PSPTLQCSPCQE 280
           |    |:|: :|          ||  ::||                    ||  |:|  |
Db      60 PALPGTGYEDVVWIP------   ---KGSVHIFIQDLNLSLSHLALKGDQFSLLLEGLPGT 109

Qy     281 HSPNSIPVVPPTPVITPQRCSPR QTPKIPCPDMADFTPKERDYAAKDCVVQTPPYQPI 339
           |     :  ||   ::| :: |: :  ||: |  |         |:   : : |  ||
Db     110 ---PQPHRLPLAGTTPHLRQGHDQAQSLEALGFINASLIIMV LAQAELPALHYRFNAPI 165

Qy     340 S--     SQWEQTDMTPCPVTCGGGVQLRSALCVDIRLKEVVTDHYCHYYFLNVEFPKL 393
           :       : |    :  |:  | || |:   || :      |  |||  : :   || :
Db     166 ASDALLPYSWHYAIWTPCSAS-DASSGLVLVYCPNQLPGSAVALRYPSDRCE  SFE SL  222

Qy     394 EECSMLPCPSSEGPKEIMPYIWFQPLPPWEHNPWTACSVSCGVGVQRRSTVCVEESMHGE 453
           : |: :|||                | |    |: || ||  |:: || ||       |
Db     223 EACNTEPCP         -       -- PSWVVCGNWSPCSRSCGLAGVESRSVVCQERVSAAE 265

Qy     454 HQVELVMPCPSASPRPVMQLTGLEAPLWAMEWSQCTVPCDEGWQVRGLRRSPVCVEESMHGE 513
           :::  |   |:| |:: |    | |:|  ::|||:|| :|| |||:|||| :
Db     266 PKAIDNSACP-QPPPPVLPAGQCPMCPPPWATLWSPSCTPSCDGLRHPVVLCKS-      319

Qy     514 VCQCNPQLELHTPFECVTP- --TPCYPCEESCVVAPLPWELQDAQELEELTKTAIEEPTF 568
                   :    :|    :|  ||   | :   |       :     | | |
Db     320            ADQRSTLTTCHCLPAAKP  PSTMRCNL-  -    RRCPPARWVTSF  - 357

Qy     569 IPEPWSACSITCGPGVQERVVECRVITGQIEELPEEECEGPELPTE- RPCLLEACD 626
           |  ||| || | | | |:|        : | | ||      || ||  :: |:
Db     358 ----WGECSTQCGLGQQQRTVRC-------ISHTGQPSREC-     -IEAERPSTMQQCE 400

Qy     627 ESPACRELEDFIFRNGSETYRWETASFT                  CIATTLV 662
           |  :  :|  :  |   |      : |                  |  || |
Db     401     AN LAVVLPGGLEECFRVNEVAT-TYVLPLGPLEAV-P_MTPFSVG 449


RESULT  11
ATS6_HUMAN
ID   ATS6_HUMAN     STANDARD;      PRT;   860 AA.
AC   Q9UKP5;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   ADAMTS 6 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE   with thrombospondin motifs 6) (ADAM-TS 6) (ADAM-TS6).
GN   ADAMTS6.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99395124; PubMed=10464288;
RA   Nakayama T.T., Hirohata S., Seldin M.F., Apte S.S.;
RT   "ADAM-TS5, ADAM-TS6, and ADAM-TS7, novel members of a new family of
RT   zinc metalloproteases.";
RL   J. Biol. Chem. 274:25555-25563(1999).
CC   -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: SECRETED, ASSOCIATED WITH THE EXTRACELLULAR
CC       MATRIX (BY SIMILARITY).
CC   -!- TISSUE SPECIFICITY: EXPRESSED AT LOW LEVELS IN PLACENTA AND BARELY
CC       DETECTABLE IN A NUMBER OF OTHER TISSUES.
CC   -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE 1 DOMAINS ARE IMPORTANT
CC       FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC   -!- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY
CC       SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC   -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 2 TSP TYPE 1 DOMAINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

DR EMBL; AF140674; AAD56522.1; -.
DR HSSP; P41279; 1IAG.
DR MIM; 605008;
DR InterPro; IPR001762; Disintegrin.
DR InterPro; IPR002870; Pep_M12B_propep.
DR InterPro; IPR001590; Peptidase_M.
DR InterPro; IPR000084; ISPI.
DR InterPro; IPR000130; Zn_MIpeptidase.
DR Pfam; PF01562; Pep_M12B_propep; 1.
DR Pfam; PF01421; Reprolysin; 1.
DR Pfam; PF00090; tsp_1; 1.
DR SMART; SM00209; ISPI; 1.
DR PROSITE; PS50215; ADAM_MEPRO; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
DR PROSITE; PS50092; ISPI; 1.
DR PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Repeat; Extracellular matrix.
FT SIGNAL 1 21 POTENTIAL.
FT PROPEP 22 244 BY SIMILARITY.
FT CHAIN 245 860 ADAM15.
FT METAL 403 403 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 404 404 BY SIMILARITY.
FT METAL 407 407 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 413 413 ZINC (CATALYTIC) (BY SIMILARITY).
FT DOMAIN 453 514 DISINTEGRIN LIKE.
FT DOMAIN 510 566 ISP TYPE I 1.
FT DOMAIN 567 628 CYS RICH.
FT DOMAIN 629 715 SPACER.
FT DOMAIN 796 852 ISP TYPE I 2.
FT DOMAIN 68 71 POLY ARG.
FT DOMAIN 662 665 POLY GLY.
FT CARBOHYD 99 99 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 172 172 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 222 222 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 676 676 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 843 843 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 860 AA; 97106 MW; E57210F0DE0B6A5C CRC64;

Query Match 5.7%; Score 520.5; DB 1; Length 860;
Best Local Similarity 29.1%; Pred. No. 5.4e-26;
Matches 125; Conservative 56; Mismatches 144; Indels 111; Gaps 13;

Qy 75 DCNWSAWCFWSCWCDKSKTCQSLAMYCHPPYLG 108
Db 510 DCGWCDWSWCWCPCSKTCQGSVSSCLPHPCDGPAFPQCCKYCLGEPKPYSCNCIDFPCDSP 569

Qy 109 DFRAQGCAYNDVQYQCHYYEWLPFYNDPAAINALKCHACQCNEVVELAPKVLGTPCNC 166
Db 570 EFKEPGTANTCNMFDCKYYNWKFYCCIVKECAQLNCLAESTNETLEKAVAVICLPGNA 629

Qy 167 DSLDMCIDGFCQAVDCRQDCNAKEINCDVCACDSGFCKHVSGFKKLEENVT 218
Db 630 DSLDGCIDNGDCKHVSCNTGSSLAKECRCKVGSSGSDCQATDSGFENLSL PKSGYMEVV 689

Qy 219 AVPLGSKSVRLTVKCPAHLETGKTLQCS KGEHSFNSDCVIVVENIVDSGSS E 274
Db 690 QIPPGSVHLTVREVAMNFKNVACKS PGEGVVINCAWTFWPPKEFVACTAFDVEPTGTT 749

Qy 275 KQTFKTDCPELMADGTLETKTKYGAAKCGVVQETTYQPTSHQWWQTDFPQCTVTSTSYQDNF 334
Db 750 PESLEALGPTSENLIVMV IIAGGNIATIRYKFNVPT IRESSGNEV 799

Qy 335 AFCVDTKLKPVVTDGVCHYYFDNVSPPKFKLKDICMECTQSGIQCKKSIVHEGLTCWWL 394
Db 800 GF SWN 860

Qy 424 HNFWTACSVSCGGG - GGEFR VCVLLGMRDLTLQVLEWSMYAPK 467
Db 798 HQPWSECSATCAGGKMRCRQPTGPARWRTKHILSYAICLEKKLDNI SCEFA 849

Qy 468 PKVMQTCNL 476
Db 850 SSCNL 858

RESULT 12
AIS8_HUMAN
ID AIS8_HUMAN STANDARD; PRT; 860 AA.
AC Q00P79; Q9NZS0;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE ADAM15 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 8) (ADAM-15) (ADAM 15) (METH-2)
DE (METH 8).
GN ADAM15 OR METH2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RX MEDLINE=99367466; PubMed=10438512;
RA Vazquez F., Hastings G., Ortega M.-A., Lane T.F., Oikemus S.,
RA Lombardo M., Iruela-Arispe M.-L.;
RT "METH-1, a human ortholog of ADAMTS-1, and METH-2 are members of a new
RT family of proteins with angio-inhibitory activity.".
RL J. Biol. Chem. 274:23349-23357(1999).
RN [2]
RP SEQUENCE OF 195-440 FROM N.A.
RX MEDLINE=20079168; PubMed=10871729;
RA Georgiadis K.E., Hirohata S., Seldin M.F., Apte S.S.;
RT "ADAM-15L, a novel metalloprotease of the ADAM-15 family located on
RT mouse chromosome 9 and human chromosome 11.";
RL Genomics 62:412-415(1999).
CC -!- FUNCTION: HAS ANTI-ANGIOGENIC PROPERTIES.
CC -!- CATALYTIC ACTIVITY: ZINC ION (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: SECRETED, ASSOCIATED WITH THE EXTRACELLULAR
CC MATRIX (BY SIMILARITY).
CC -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN ADULT AND FETAL LUNG, LOWER
CC EXPRESSION IN BRAIN, PLACENTA, HEART, SKELETAL MUSCLE AND FETAL BRAIN AND
CC KIDNEY.
CC -!- DOMAIN: THE SPACER DOMAIN AND THE ISP TYPE I DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -!- PTM: THE PRO-FORM IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY
CC SIMILARITY).
CC -!- SIMILARITY: BELONGS TO 1 PEPTIDASE FAMILY M12B.
CC -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -!- SIMILARITY: CONTAINS 2 ISP TYPE I DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AF092254; AAD19081.1; -.
DR EMBL; AF175283; AAF25806.1; -.
DR HSSP; P41279; 1IAG.
DR MIM; 605175; -.
DR InterPro; IPR001762; Disintegrin.
DR InterPro; IPR002870; Pep_M12B_propep.
DR InterPro; IPR001590; Peptidase_M.
DR InterPro; IPR000084; ISPI.
DR InterPro; IPR000130; Zn_MIpeptidase.

```
DR   Pfam; PF01562; Rep_M12B_propep; 1.
DR   Pfam; PF01421; Reprolysin; 1.
DR   Pfam; PF00090; tsp_1; 2.
DR   SMART; SM00209; TSP1; 2.
DR   PROSITE; PS50215; ADAM_MEPRO; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
DR   PROSITE; PS50092; TSP1; 1.
DR   PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
KW   Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW   Repeat; Extracellular matrix; Heparin-binding.
FT   SIGNAL        1     27       POTENTIAL.
FT   PROPEP       28    214       BY SIMILARITY.
FT   CHAIN       215    890       ADAMTS-8.
FT   METAL       364    364       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    365    365       BY SIMILARITY.
FT   METAL       368    368       ZINC (CATALYTIC) (BY SIMILARITY).
FT   METAL       374    374       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DOMAIN      439    526       DISINTEGRIN LIKE.
FT   DOMAIN      527    583       TSP TYPE 1 1.
FT   DOMAIN      584    690       CYS-RICH.
FT   DOMAIN      691    832       SPACER.
FT   DOMAIN      844    896       TSP TYPE 1 2.
FT   DOMAIN      202    255       BY SIM.
FT   CARBOHYD    345    345       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    431    431       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    466    466       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    491    491       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    606    606       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT    195    195       E -> R (IN REF. 2).
FT   CONFLICT    418    440       YLTELLDGGHGDGLLDAPGAALPLPTGL -> FSGCHLQGW
FT                                 IHPKYLCKCVSELKCTLMP (IN REF. 2).
SQ   SEQUENCE   890 AA;  96671 MW;  57D70EE03D57C0E4 CRC64;
```

```
  Qy, Match              5.3%;  Score 196.5;  DB 1;  Length 890;
  Best Local Similarity   31.9%;  Pred. No. 7.1e-24;
  Matches  122;  Conservative   55;  Mismatches  131;  Indels   75;  Gaps   19;

Qy      75 DGNWDAWGDWSDCSRTCGGGASYSLRPCLT------GRNCPGQNIRYRTCSNPDCPPDAF 128
           || |  || | :|||||||  :| | |        || | |:  :|:::|  :|||| :
Db     527 DGGWAPWGEWGECSRTCGGGVQFSHRECKDPEPQNGGRYCLGPRAKYQSCHIEDCPPDGK 586

Qy     129 DFRAQGCSAYNEVQY   QGHYYFWLPRYN- DFAAPGAEKCHAGG-NLVVELAPKVLLG 183
           || |||  ||  :|    |:  :|:|:|    |  | | | |:|::       |||:|
Db     587 SFRPQQCPKYNAYNYIPMGCNIIQWVPKYAGVSPRDPMPLPCRAPCDGPPPVPEAKVIDC 646

Qy     184 TFCNIEYIMMCNGCIGAVCDDD-DKCNAPTLNCIVGAP SPCPFVKLPFSHVVTPFS 239
           | |  :|| :|: | |  ||| : |  | | |||| | |:|| | |     :||
Db     647 TLCGPETLAICVRGQCVPAGIDHVVESPRKLEHGSVCGCPCNSQPPVSPE-  TPTTNY 702

Qy     240 EEN VIAVELGSRSVKLTVKCPAHLPTDCK    TLGYCPLDLPSFNG       HGVLVV 291
           | :: :| |  :  | ||  :|  :::      |: : |::  |           :|
Db     703 GYNDIVTIPAG--ATNIDVKQKSHPGVQNDGNYLALKTADGQYTLNGNLAIGATEQDILV 760

Qy     292 ENTTVEP-S-C    PL TPP II YIMAGIIF        RTPTIMAAFLVV_IF     -4
           : | :::  ||    :| |:::: | ||    :         | :||    : | |
Db     761 KGIILKYS GNIATIPRL-SPRPIPPPITV_IEIVPSEVFIIPRVKYIIEVENDVYHSMQV 819

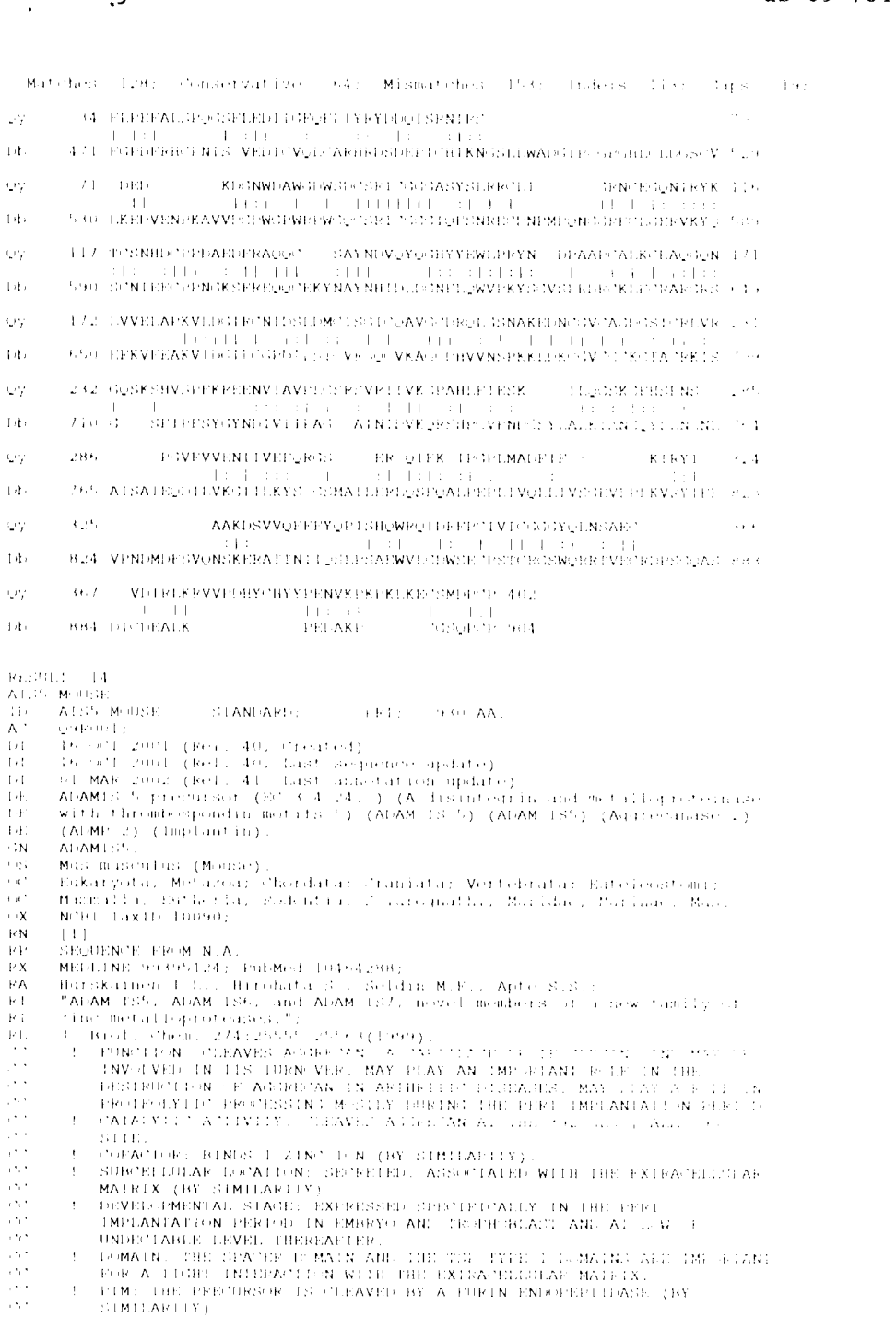
Qy     335 --------FYQPISH-QWPQTDFFPCTVTGGGGY_LNSAICVLIKLRPVVFTHY    GH 381
                  || : ||   |:  : :|| :|:  : || |       |       |:
Db     820 SKEPATTNII_PLIHA_WVLQDWSPCSSTCGAGWQRRTVECRD-- -  -PSQQAEATCN 878

Qy     382 --YYPENVKPKPKLAGSMHPLP 402
           |: :|     |   ||
Db     879 KALKPEDAKP -----GLSQLCP 899
```

```
RESULT  14
AIS8_MOUSE
ID   AIS8_MOUSE     STANDARD;      PRT;   905 AA.
AC   P57110;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
```

```
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   ADAMTS-8 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE   with thrombospondin motifs 8) (ADAM-TS 8) (ADAM-TS8) (METH-2).
GN   ADAMTS8.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20079168; PubMed=10610729;
RA   Georgiadis K.E., Hirohata S., Seldin M.F., Apte S.S.;
RT   "ADAM-TS8, a novel metalloprotease of the ADAM-TS family located on
RT   mouse chromosome 9 and human chromosome 11.";
RL   Genomics 62:312-315(1999).
CC   -!- FUNCTION: HAS ANTI ANGIOGENIC PROPERTIES (BY SIMILARITY).
CC   -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: SECRETED, ASSOCIATED WITH THE EXTRACELLULAR
CC       MATRIX (BY SIMILARITY).
CC   -!- TISSUE SPECIFICITY: EXPRESSED SPECIFICALLY IN ADULT LUNG AND HEART
CC       AND LOW EXPRESSION DURING MOUSE DEVELOPMENT.
CC   -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE 1 DOMAINS ARE IMPORTANT
CC       FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC   -!- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY
CC       SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC   -!- SIMILARITY: CONTAINS 1 DISINTEGRIN LIKE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 2 TSP TYPE 1 DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF175282; AAF25805.1; -.
DR   MGD; MGI:1353468; Adamts8.
DR   InterPro; IPR001762; Disintegrin.
DR   InterPro; IPR001590; Reprolysin.
DR   InterPro; IPR000884; TSP1.
DR   InterPro; IPR000130; Zn_MEpeptdse.
DR   Pfam; PF01421; Reprolysin; 1.
DR   Pfam; PF00090; tsp_1; 2.
DR   SMART; SM00209; TSP1; 2.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
DR   PROSITE; PS50215; ADAM_MEPRO; 1.
DR   PROSITE; PS50092; TSP1; 1.
DR   PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
KW   Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW   Repeat; Extracellular matrix; Heparin-binding.
FT   SIGNAL        1     28       POTENTIAL.
FT   PROPEP       29    229       BY SIMILARITY.
FT   CHAIN       230    905       ADAMTS-8.
FT   METAL       378    378       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    379    379       BY SIMILARITY.
FT   METAL       382    382       ZINC (CATALYTIC) (BY SIMILARITY).
FT   METAL       388    388       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DOMAIN      453    541       DISINTEGRIN LIKE.
FT   DOMAIN      542    598       TSP TYPE-1 1.
FT   DOMAIN      599    705       CYS-RICH.
FT   DOMAIN      706    847       SPACER.
FT   DOMAIN      848    905       TSP TYPE-1 2.
FT   CARBOHYD    415    415       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    480    480       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    506    506       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    615    615       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   905 AA;  98879 MW;  124D4132B33A0CAE CRC64;
```

```
  Query Match              5.2%;  Score 484.5;  DB 1;  Length 905;
  Best Local Similarity   27.9%;  Pred. No. 4.2e-23;
```



Matches 120; Conservative 64; Mismatches 153; Indels 116; Gaps 16;

Qy 34 FLPEPALSP...GSF...LPDL...TGDLFLIYRYLDGISKNIK 67
Db 471 PGPDPRRCNIS VPDICVGLCAPHRDSDPFDCHTKNGGLLWADGTCVPGRDDCLDGSDEV 530

Qy 71 DED KDGNWDAWSDWSRCSKDCGSGASYSLRRCLI DKNCDGLNIKYK 116
Db 531 LEEDVENPKAVVFCRWCPWPFWGSDSPSPDGQFSNNDTNPMPGNCSPFTLGPEVKVL 590

Qy 117 TCSNHDCPPDAENPRAGGC SAYNDVQYGGHYYEWLPRYN DPAAPCALKCHAQSQN 171
Db 591 SCNIEPCPPNGKSFRDGGCTEKYNAYNHILDLGNPLGWVPKYSCVSRLDKKFCRAGGS 650

Qy 172 LVVELAPKVLDGTPCNIGSLDMCISGICQAVSCDRQLRGNAKEDNCGVCAGDGSICPLVR 231
Db 651 IPKYFEAKVIDGTLCGPDKGSFCVRGQCVKAGLIGDVVNSPKLLGKCVVSKGIAYGEKS 710

Qy 232 GQSKSHVSPPKPEENVIAVELNSKPVKELVKSPAHLFIRSK ILQSPKPSNLNS 285
Db 711 G SPTPPSYGYNDGVLTFAST AINCPVFLPSDRVFNPCSYLALLSNLYLLNNSNS 764

Qy 286 PGVFVVENTLVEPQRGS ER QTEK IPGPLMADFIF KIRYI 328
Db 765 AISAIEQDGLVKGTLLKYS GSMALLEMLQGPQALPFFLIVQLLLVSVGVFLFKVSYIF 824

Qy 329 AAKDSVVQFFPYQPISHQWPQHDPPPCTVICQGGGYQLNSARS 370
Db 825 VPNDMDFSVQNSKFRAFINIQSLRSTAEWVLCHWSPCTSCQSQWQRRIVIRTPSVSQAA 884

Qy 367 VETRLKPVVPQHYCHYYPENVKDKPKLKEKVSMHYP 402
Db 885 DVCDEALK PELAKF YQAQPCP 904

RESULT 14
AL5_MOUSE
ID AL5_MOUSE STANDARD; PRT; 930 AA.
AC Q08001;
DT 15-OCT-2001 (Rel. 40, Created)
DT 15-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE ADAM15 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 7) (ADAM-15) (ADAM 15) (Metalloprotease
DE (ADAM 2) (Implantin).
GN ADAM15.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96395124; PubMed=8798298;
RA Horiuchi H.L., Hirohata S., Seldin M.F., Apte S.S.;
RT "ADAM 15S, ADAM 15L, and ADAM 15Z, novel members of a new family of
RT the metalloproteases.";
RL J. Biol. Chem. 271:29553-29558(1996).
CC -!- FUNCTION: CLEAVES AGGRECAN, A CARTILAGE PROTEOGLYCAN. MAY BE
CC INVOLVED IN ITS TURNOVER. MAY PLAY AN IMPORTANT ROLE IN THE
CC DESTRUCTION OF AGGRECAN IN ARTHRITIC DISEASES. MAY CLEAVE A
CC PROTEOLYTIC PROCESSING SITE DURING THE PERI-IMPLANTATION PERIOD.
CC -!- CATALYTIC ACTIVITY: CLEAVES ALTERNATING A-... BONDS, WHERE ... IS
CC SER.
CC -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: GENERALLY ASSOCIATED WITH THE EXTRACELLULAR
CC MATRIX (BY SIMILARITY).
CC -!- DEVELOPMENTAL STAGE: EXPRESSED SPECIFICALLY IN THE PERI-
CC IMPLANTATION PERIOD IN EMBRYO AND PERIPHERAL AND AT LOW
CC UNDETECTABLE LEVEL THEREAFTER.
CC -!- DOMAIN: THE SPACER DOMAIN AND CYS-RICH TYPE I THROMBOSPONDIN
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -!- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY
CC SIMILARITY)
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -!- SIMILARITY: CONTAINS 2 TSP TYPE 1 DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AF140673; AAD56356.1; -.
DR MGD; MGI:1346321; Adam15.
DR InterPro; IPR001762; Disintegrin.
DR InterPro; IPR000884; Pep_M12B_propep.
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000884; TSP1.
DR InterPro; IPR006143; Zn_Mlpeptdse.
DR Pfam; PF01562; Pep_M12B_propep; 1.
DR Pfam; PF01421; Reprolysin; 1.
DR Pfam; PF00090; tsp_1; 2.
DR SMART; SM00200; TSP1; 2.
DR PROSITE; PS50215; ADAM_MEPRO; 1.
DR PROSITE; PS50092; TSP1; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
DR PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Repeat; Extracellular matrix.
FT SIGNAL 1 21 POTENTIAL.
FT PROPEP 22 261 POTENTIAL.
FT CHAIN 262 930 ADAM15.
FT SITE 209 209 CYSTEINE SWITCH (POTENTIAL).
FT METAL 410 410 ZINC (CATALYTIC) (BY SIMILARITY)
FT ACT_SITE 411 411 BY SIMILARITY.
FT METAL 414 414 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 420 420 ZINC (CATALYTIC) (BY SIMILARITY).
FT DOMAIN 485 569 DISINTEGRIN-LIKE.
FT DOMAIN 567 623 TSP TYPE 1 1.
FT DOMAIN 624 731 CYS-RICH.
FT DOMAIN 732 874 SPACER.
FT DOMAIN 875 930 TSP TYPE 1 2.
FT DOMAIN 41 46 POLY ALA.
FT DOMAIN 257 261 POLY ARG.
FT CARBOHYD 408 408 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 728 728 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 802 802 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 807 807 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 930 AA; 101780 MW; 6A1D04A2C17A9A8C CRC64;

Query Match 5.2%; Score 482; DB 1; Length 930;
Best Local Similarity 26.3%; Pred. No. 6.6e-24;
Matches 134; Conservative 67; Mismatches 170; Indels 130; Gaps 21;

Qy 34 FLPEPALSP GSF LPDL TGDLFLIYRYLDGISKNIK 67
Db 471 PGPDPRRCNIS VPDICVGLCAPHRDSDPFDCHTKNGGLLWADGTCVPGRDDCLDGSDEV 530

Qy 71 DED KDGNWDAWSDWSRCSKDCGSGASYSLRRCLI DKNCDGLNIKYK 116
Db 531 LEEDVENPKAVVFCRWCPWPFWGSDSPSPDGQFSNNDTNPMPGNCSPFTLGPEVKVL 590

Qy 117 TCSNHDCPPDAENPRAGGC SAYNDVQYGGHYYEWLPRYN DPAAPCALKCHAQSQN 171
Db 591 SCNIEPCPPNGKSFRDGGCTEKYNAYNHILDLGNPLGWVPKYSCVSRLDKKFCRAGGS 650

Qy 172 LVVELAPKVLDGTPCNIGSLDMCISGICQAVSCDRQLRGNAKEDNCGVCAGDGSICPLVR 231
Db 651 IPKYFEAKVIDGTLCGPDKGSFCVRGQCVKAGLIGDVVNSPKLLGKCVVSKGIAYGEKS 710

Qy 232 GQSKSHVSPPKPEENVIAVELNSKPVKELVKSPAHLFIRSK ILQSPKPSNLNS 285
Db 711 GSPTPPSYGYNDGVLTFASTAINCPVFLPSDRVFNPCSYLALLSNLYLLNNSNS 764

Qy 286 PGVFVVENTLVEPQRGS ER QTEK IPGPLMADFIF KIRYI 328
Db 765 AISAIEQDGLVKGTLLKYSGSMALLEMLQGPQALPFFLIVQLLLVSVGVFLFKVSYIF 824

Qy 329 AAKDSVVQFFPYQPISHQWPQHDPPPCTVICQGGGYQLNSARS 370
Db 825 VPNDMDFSVQNSKFRAFINIQSLRSTAEWVLCHWSPCTSCQSQWQRRIVIRTPSVSQAA 884

Qy 367 VETRLKPVVPQHYCHYYPENVKDKPKLKEKVSMHYP 402
Db 885 DVCDEALK PELAKF YQAQPCP 904

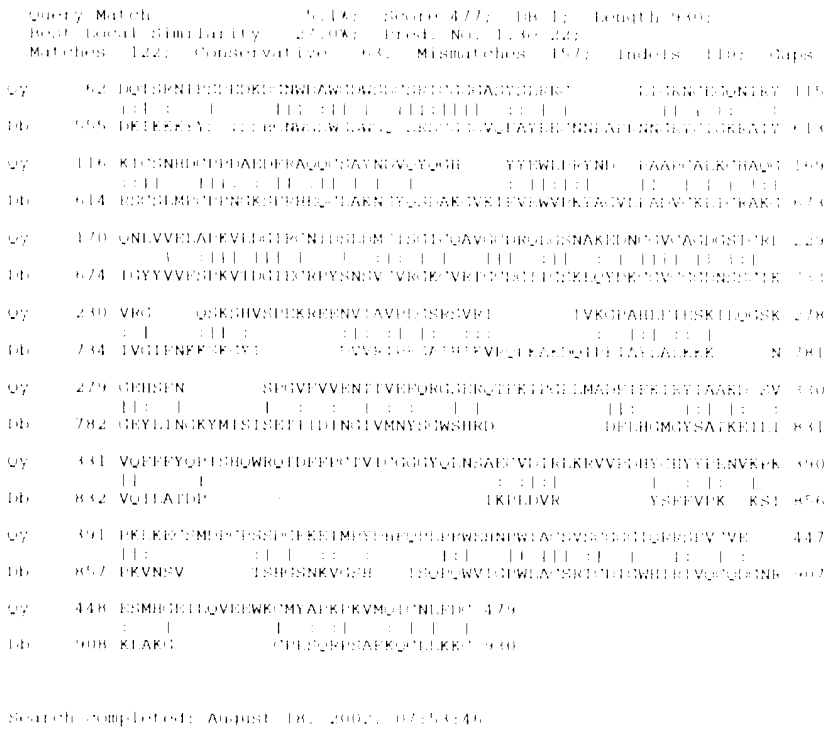
```
Db    740 FFSPGTT    -    SVVEIIEDATHIEVEQEFAPPQIEIAVIAEEFI      781 IN 787

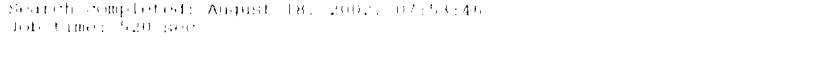
Qy    345 -------SPGVPVVPNTIVPPQVGSNPQTPPIPGDIMANMIPVTPVIAAVD-FVVQSTPV 396
                  |  :   :  | : :   | |           ||:    :|  |:  :||
Db    788 GKYMISISGETIIDINGTVMNYSGWSHRD- ---     DFLRGMGYSATKEILIVQILAT 837

Qy    397 SMDPGRSSIWFKFIMPVTHPPQPIDPWFHNPWIACSVSCGGGIQRPSIVCVR---ESMHQE 453
                  | |   :: |  |   | :|   || ||| :|  |   |:  | :   :   |
Db    861 ----SVISHGSNKVGP--HSTQL-QWVIGPWIACSRTCDTGWHTPTVQCQDGNPFIAEG- 912

Qy    454 IIQVEPWPCMYAPEPVMQITNLEC 477
                  |: : :|   : | |  |
Db    913 --------CLISQRPSAFKQCLIKK 930


RESULT 15
AIS5_HUMAN
ID   AIS5_HUMAN     STANDARD;      PRT;   930 AA.
AC   Q9UNA0; Q9UKP2;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   01 MAR 2002 (Rel. 41, Last annotation update)
DE   ADAMTS-5 precursor (EC 3.4.24. ) (A disintegrin and metalloproteinase
DE   with thrombospondin motifs 5) (ADAM TS 5) (ADAM TS5) (Aggrecanase-2)
DE   (ADMP-2) (ADAM-TS 11).
GN   ADAMTS5 OR ADMP2 OR ADAMTS11.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=99367476; PubMed=10400522;
RA   Abbaszade I., Liu R.Q., Yang F., Rosenfeld S.A., Ross O.H.,
RA   Link J.R., Ellis D.M., Tortorella M.D., Pratta M.A., Hollis J.M.,
RA   Wynn R., Duke J.L., George H.J., Hillman M.C. Jr., Murphy K.,
RA   Wiswall B.H., Copeland R.A., Decicco C.P., Bruckner K., Nagase H.,
RA   Ito Y., Newton R.C., Magolda R.L., Trzaskos J.M., Hollis G.F.,
RA   Arner E.C., Burn T.C.;
RT   "Cloning and characterization of ADAMTS11, an aggrecanase from the
RT   ADAMTS family.";
RL   J. Biol. Chem. 274:23443-23450(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20289799; PubMed=10830953;
RA   Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
RA   Park H.S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,
RA   Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K.,
RA   Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
RA   Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
RA   Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
RA   Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
RA   Minoshima S., Shimizu N., Nordsiek G., Hornischer K., Brandt P.,
RA   Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloecker H.,
RA   Ramser J., Beck A., Klages S., Hennig S., Riesselmann L., Dagand E.,
RA   Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
RA   Lehrach H., Reinhardt R., Yaspo M.-L.;
RT   "The DNA sequence of human chromosome 21.";
RL   Nature 405:311-319(2000).
RN   [3]
RP   SEQUENCE OF 413-930 FROM N.A.
RC   TISSUE=Fetal brain;
RX   MEDLINE=99295124; PubMed=10464298;
RA   Hurskainen T.L., Hirohata S., Seldin M.F., Apte S.S.;
RT   "ADAM-TS5, ADAM-TS6, and ADAM-TS7, novel members of a new family of
RT   zinc metalloproteases.";
RL   J. Biol. Chem. 274:25555-25563(1999).
CC   -!- FUNCTION: CLEAVES AGGRECAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE
CC       INVOLVED IN ITS TURNOVER. MAY PLAY AN IMPORTANT ROLE IN THE
CC       DESTRUCTION OF AGGRECAN IN ARTHRITIC DISEASES. MAY PLAY A ROLE IN
CC       PROTEOLYTIC PROCESSING MOSTLY DURING THE PERI-IMPLANTATION PERIOD.
CC   -!- CATALYTIC ACTIVITY: CLEAVES AGGRECAN AT THE 392-GLU-|-ALA-393
CC       SITE.
CC   -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR
CC       MATRIX (BY SIMILARITY).
CC   -!- TISSUE SPECIFICITY: EXPRESSED AT LOW LEVEL, PRIMARILY IN PLACENTA
CC       BUT ALSO IN OTHER TISSUES, SUCH AS HEART AND BRAIN, AND ALSO
CC       CERVIX, UTERUS, BLADDER, ESOPHAGUS, RIB CARTILAGE,
CC       CHONDROBLASTOMA, FIBROUS TISSUE AND JOINT CAPSULE FROM AN
CC       ARTHRITIC PATIENT.
CC   -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
CC       FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC   -!- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY
CC       SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC   -!- SIMILARITY: CONTAINS 1 DISINTEGRIN LIKE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 2 TSP TYPE-1 DOMAINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AF212099; AAD19577.1; -.
DR   EMBL; AP001698; BAA95504.1; -.
DR   EMBL; AP001697; BAA95503.1; -.
DR   EMBL; AF141293; AAF92497.1; -.
DR   HSSP; Q9PW35; 1BUD.
DR   MIM; 605007; -.
DR   InterPro; IPR001762; Disintegrin.
DR   InterPro; IPR001590; Reprolysin.
DR   InterPro; IPR000884; TSP1.
DR   InterPro; IPR000130; Zn_MIpeptdse.
DR   Pfam; PF01421; Reprolysin; 1.
DR   Pfam; PF00090; tsp_1; 2.
DR   SMART; SM00209; TSP1; 2.
DR   PROSITE; PS50215; ADAM_MEPRO; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
DR   PROSITE; PS50092; TSP1; 1.
DR   PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
KW   Hydrolase, Metalloprotease, Zinc, Signal, Glycoprotein, Zymogen,
KW   Repeat; Extracellular matrix.
FT   SIGNAL       1     16       POTENTIAL.
FT   PROPEP      17    261       POTENTIAL.
FT   CHAIN      262    930       ADAMTS-5.
FT   SITE       209    209       CYSTEINE SWITCH (POTENTIAL).
FT   METAL      410    410       ZINC (CATALYTIC) (BY SIMILARITY)
FT   ACT_SITE   411    411       BY SIMILARITY.
FT   METAL      414    414       ZINC (CATALYTIC) (BY SIMILARITY).
FT   METAL      420    420       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DOMAIN     195    566       DISINTEGRIN LIKE.
FT   DOMAIN     567    625       TSP TYPE 1 1.
FT   DOMAIN     624    731       CYS-RICH.
FT   DOMAIN     732    874       SPACER.
FT   DOMAIN     875    930       TSP TYPE 1 2.
FT   DOMAIN      37     41       POLY-ALA.
FT   DOMAIN     257    261       POLY-ARG.
FT   CARBOHYD   498    498       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   728    728       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   802    802       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   807    807       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT   138    138       A -> G (IN REF. 2).
FT   CONFLICT   614    614       R -> H (IN REF. 3).
FT   CONFLICT   692    692       P -> L (IN REF. 2).
SQ   SEQUENCE   930 AA;  101715 MW;  B64281502F28193B CRC64;
```


Query Match 5.1%; Score 477; DB 1; Length 930;
Best Local Similarity 27.0%; Pred. No. 1.3e-22;
Matches 122; Conservative 63; Mismatches 197; Indels 110; Gaps 18;



Search completed: August 18, 2002, 07:54:46
Job time: 520 sec

Geo ... ce Version 4.1
Copyright (c) ... - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: ... Mon August 19, 2002, 5:15:41 ; Search Time 49.1 Seconds
(without alignments)
4574.289 Million cell updates/sec

Title: US-09-784-358-2
Perfect score: 9446
Sequence: 1 MASWTSPWWVFIGMVFMHSP.....ENLVSLIHFYKLRDVLSQLPS 1063

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283148 seqs, 96080444 residues

Total number of hits satisfying chosen parameters: 283148

Minimum DB seq length: 0
Maximum DB seq length: 200300000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1184.5	12.7	2165	2	I21471	hypothetical prote
2	1110.5	11.8	1050	2	I29545	hypothetical prote
3	866	9.2	1558	2	S80114	protein C54G8.4 [
4	863	9.2	2167	2	I34495	hypothetical prote
5	700	7.5	962	2	I40966	hypothetical prote
6	594.5	6.4	1264	2	I29617	probable N-ethyl
7	593	6.3	858	2	I14764	hypothetical prote
8	571.5	6.1	959	2	I47158	hypothetical prote
9	564.5	6.0	951	2	I00017	gene AD/AMIS 1 prot
10	400	5.2	1444	2	I28856	amelioidosis inhib
11	415	4.4	867	2	I69455	hypothetical prote
12	452	4.8	957	2	I15976	hypothetical prote
13	441.5	4.7	788	2	I25061	hypothetical prote
14	292.5	3.1	1522	2	I20028	brain specific ang
15	289.5	3.1	1584	2	I00026	brain specific ang
16	276	3.0	803	2	A47723	F-spondin precurso
17	274	2.9	447	2	S05478	properdin - mouse
18	273	2.9	469	1	S20129	properdin precurso
19	267	2.8	984	2	I00429	hypothetical prote
20	265	2.8	867	2	A39252	F-spondin ... rat
21	261	2.8	5129	2	I20092	hypothetical prote
22	257.5	2.8	1074	2	S75928	semaphorin F precu
23	256.5	2.7	3707	2	S10252	heparan sulfate pr
24	254.5	2.7	5198	2	I43290	hemicentin precurs
25	253.5	2.7	1360	2	I38922	hypothetical prote
26	253	2.7	1496	2	I00851	beta-synaptoglycan
27	245.5	2.6	1572	2	I00027	brain specific ang
28	247	2.6	712	2	A45698	immunoglobulin mu
29	242.5	2.6	1138	1	A39804	thrombospondin pre
30	241.5	2.6	4262	2	I42533	connectin/titin
31	241	2.6	5376	2	I42215	cytadhesin - mouse
32	240	2.6	1170	1	ISHUP1	thrombospondin 1 p
33	240	2.6	1172	1	ISHUP2	thrombospondin 2 p
34	237.5	2.5	4391	2	A38096	perlecan precursor
35	237.5	2.5	7962	2	I38446	elastic titin - hu
36	236.5	2.5	724	2	A48669	antigen Ber-EP4 - H
37	235	2.5	1172	2	A42587	thrombospondin 2 p
38	234	2.5	1170	2	A40558	thrombospondin 1 p
39	233	2.5	805	2	I34212	hypothetical prote
40	230.5	2.5	1344	2	I14416	rig i protein - mo
41	229.5	2.5	6642	2	I29757	protein UNC-89 - C
42	229.5	2.4	1188	2	I86256	protein F14N23.5 [
43	215	2.3	860	2	I16892	hypothetical prote
44	215.5	2.3	1033	2	S19247	cell adhesion prot
45	215.5	2.3	1272	2	S26180	neurotactin - chic

ALIGNMENTS

RESULT 1
I21471
hypothetical protein F25H8.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C;Accession: I21471; I24896
R;Gajadsty, S.
submitted to the EMBL Data Library, February 1996
A;Reference number: Z19415
A;Accession: I21471
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2165 <WIL>
A;Cross-references: EMBL:Z69360; PIDN:CAA93287.1; GSPDB:GN00022; CESP:F25H8.3
A;Experimental source: clone F25H8
R;Gajadsty, S.
submitted to the EMBL Data Library, February 1996
A;Reference number: Z19949
A;Accession: I24896
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2165 <WIL>
A;Cross-references: EMBL:Z69361; PIDN:CAA93288.1; GSPDB:GN00022; CESP:F25H8.3
A;Experimental source: clone H10D9
C;Genetics:
A;Gene: CESP:F25H8.3
A;Map position: 4
A;Introns: 57/3 113/3 [illegible]

Query Match 12.7%; Score 1184.5; DB 2; Length 2165;
Best Local Similarity 22.7%; Pred. No. 5.4e-62;
Matches 402; Conservative 212; Mismatches 570; Indels 580; Gaps 80;

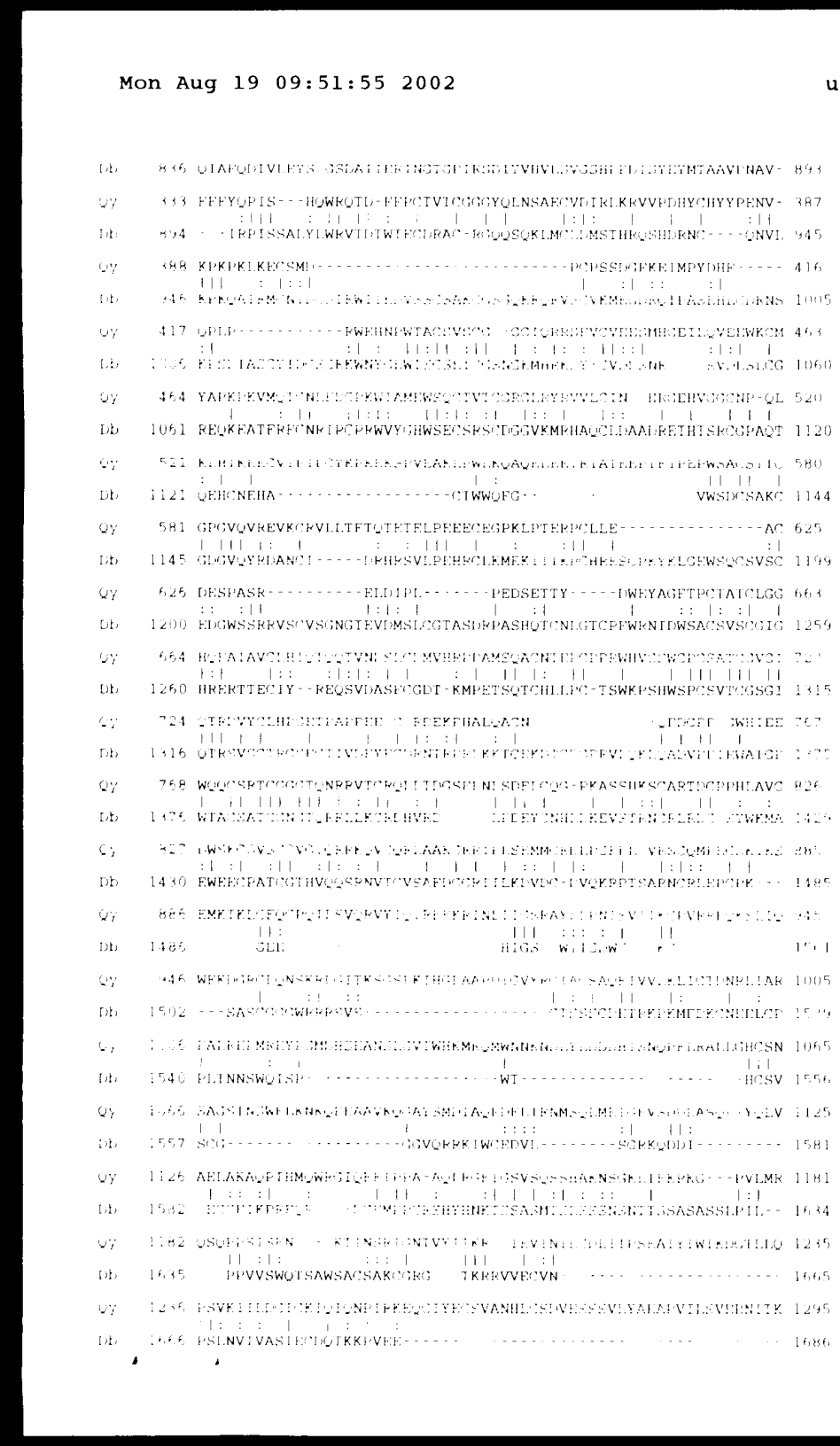
Qy 75 [illegible] 128
Db 603 [illegible] 662

Qy 129 [illegible] 178
Db 663 [illegible] 719

Qy 179 [illegible] 238
Db 720 [illegible] 779

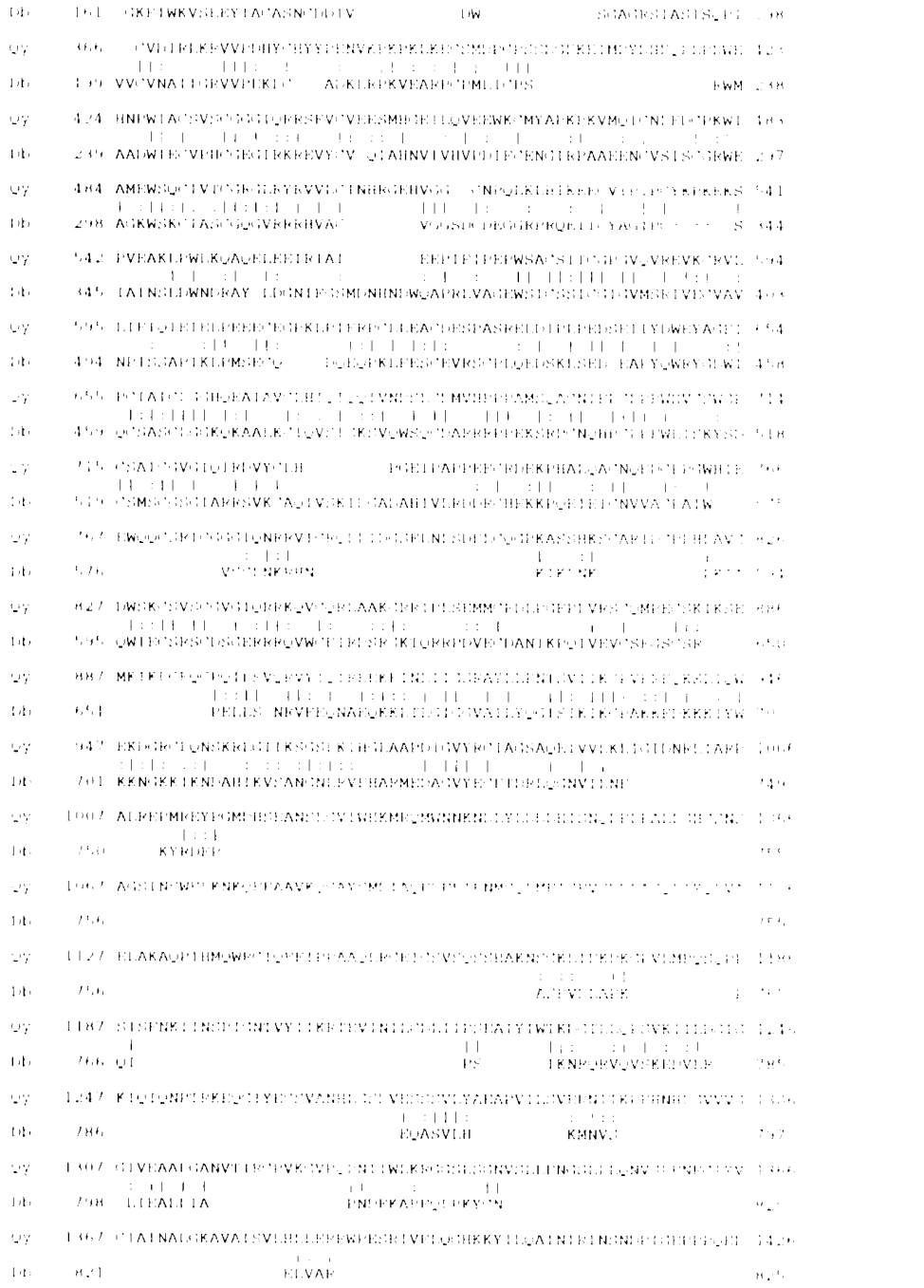
Qy 239 [illegible] 291
Db 780 [illegible] 835

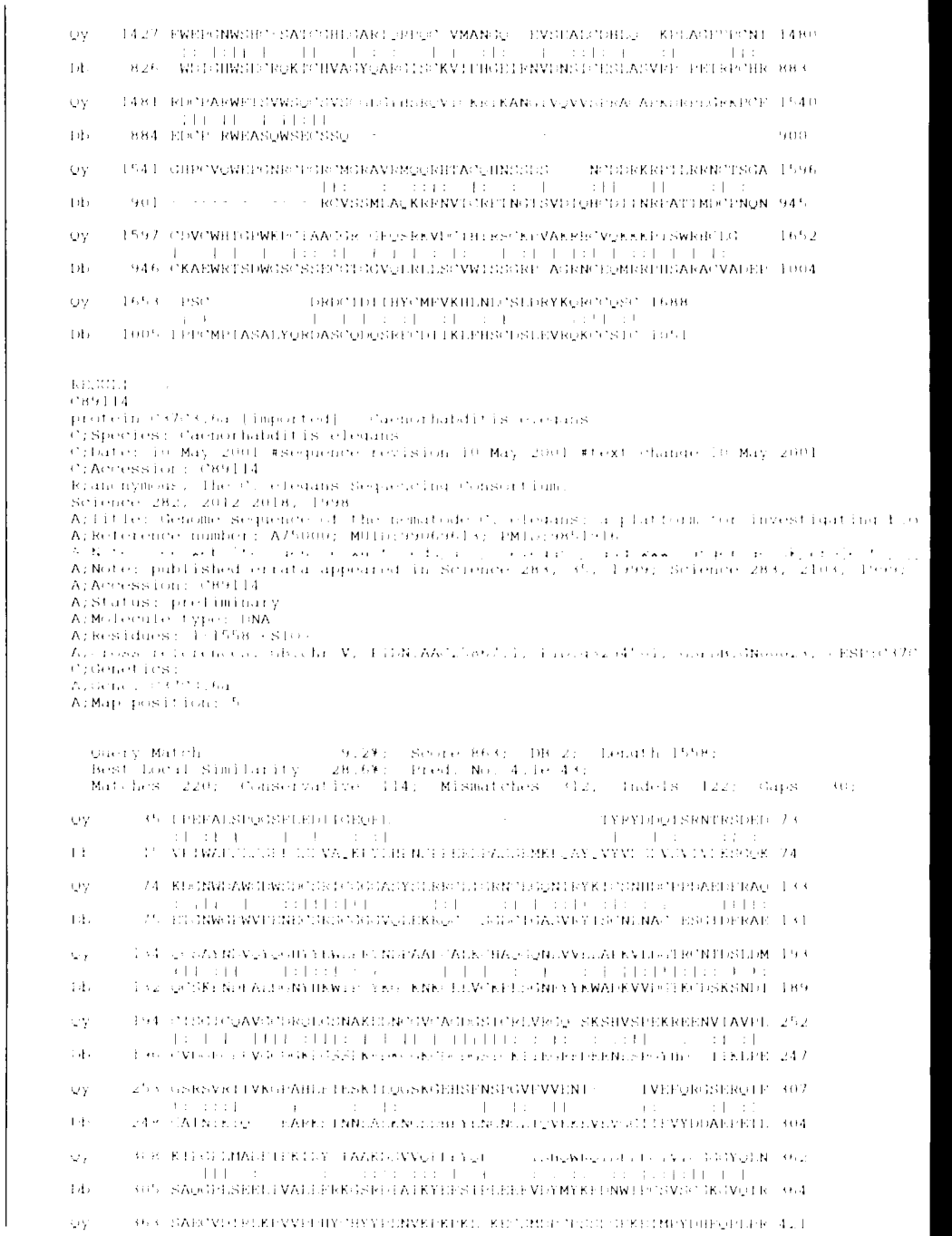
Qy 292 [illegible] 342
Db 836 [illegible] 885

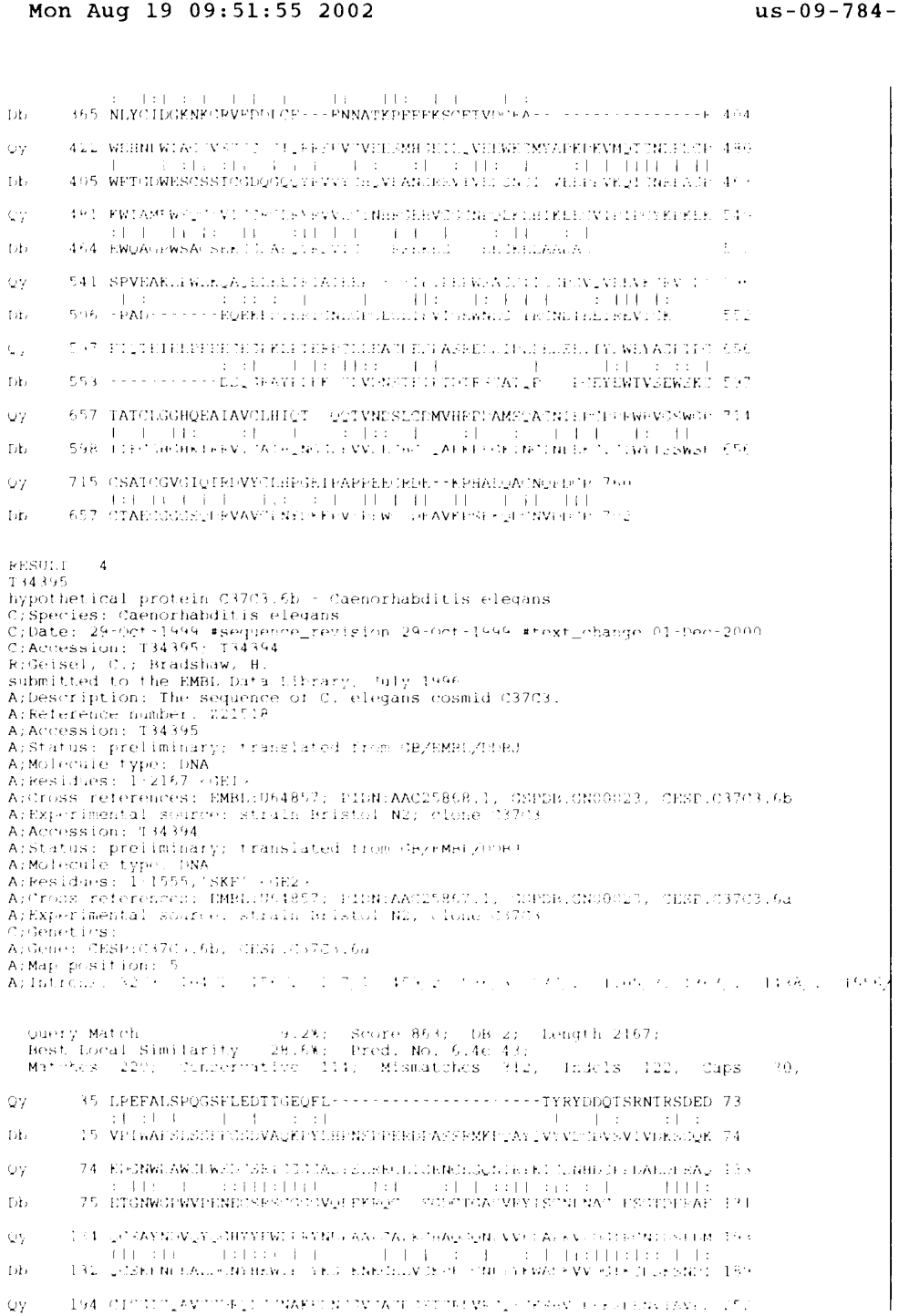


RESULT 2
T22545
hypothetical protein F53B6.2 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Accession: T22545
R;White, S.
submitted to the EMBL Data Library, October 1996
A;Reference number: 219578
A;Accession: T22545
A;Molecule type: DNA
A;Cross references: EMBL:Z81086; PIDN:CAB03121.1; GSPDB:GN00019; CESP:F53B6.2
A;Experimental source: clone F53B6
C;Genetics:
A;Map position: I
A;Introns: 88/3; 162/3; 214/3; 264/1; 298/2; 449/2; 475/2; 1...

Query Match 11.9%; Score 1110.5; DB 2; Length 1059;
Best Local Similarity 21.1%; Pred. No. 4.5e-58;
Matches 351; Conservative 170; Mismatches 451; Indels 695; Gaps 52;







RESULT 4
T34395
hypothetical protein C37C3.6b - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 01-Dec-2000
C;Accession: T34395; T34394
R;Geisel, C.; Bradshaw, H.
submitted to the EMBL Data Library, July 1996
A;Description: The sequence of C. elegans cosmid C37C3.
A;Reference number: Z22518
A;Accession: T34395
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2167 <GE1>
A;Cross-references: EMBL:U64852; PIDN:AAC25868.1; GSPDB:GN00023; CESP:C37C3.6b
A;Experimental source: strain Bristol N2; clone C37C3
A;Accession: T34394
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1555,'SKF' <GE2>
A;Cross-references: EMBL:U64852; PIDN:AAC25867.1; GSPDB:GN00023; CESP:C37C3.6a
A;Experimental source: strain Bristol N2; clone C37C3
C;Genetics:
A;Gene: CESP:C37C3.6b, CESP:C37C3.6a
A;Map position: 5
A;Introns: 52-[illegible]; 175-[illegible]; 454-[illegible]; 1105-[illegible]; 1148-[illegible]; 1695-[illegible]

Query Match 9.2%; Score 863; DB 2; Length 2167;
Best Local Similarity 28.6%; Pred. No. 6.4e-43;
Matches 220; Conservative 111; Mismatches 312; Indels 122; Gaps 20;

Qy 35 LPEFALSPQGSFLEDTTGEQFL---------------------TYRYDDQTSRNTIRSDED 73
Db 15 VPIWAFSSLSGFTGSLVAGFFVLIIFNVFFFFFRHRAYFFMKFLGAIVSVFVFVVVSVIVIKKSQK 74

Qy 74 EGGNWGLAWCLWGGVGGLTSSAGLGGLGLLGGNGNGGGGGGKGLGNHGGLGLDALGLGGAG 133
Db 75 DTGNWGFWVFFNEGSFSFSVSSVWLFFWGF--SSWTFGAFVFYISSMLNAT FSSTMFFAF 131

Qy 134 GGAYNGVGYGGHYYFWFFGVNGLAAGFAGAGGNGLVVFIAGVVGGGVNTLSLLM 193
Db 132 GGSFLNFLAGGFGNHFWGL SGLFNGGLLVFSFL SNLSFFWASFVVSFLSFSG 189

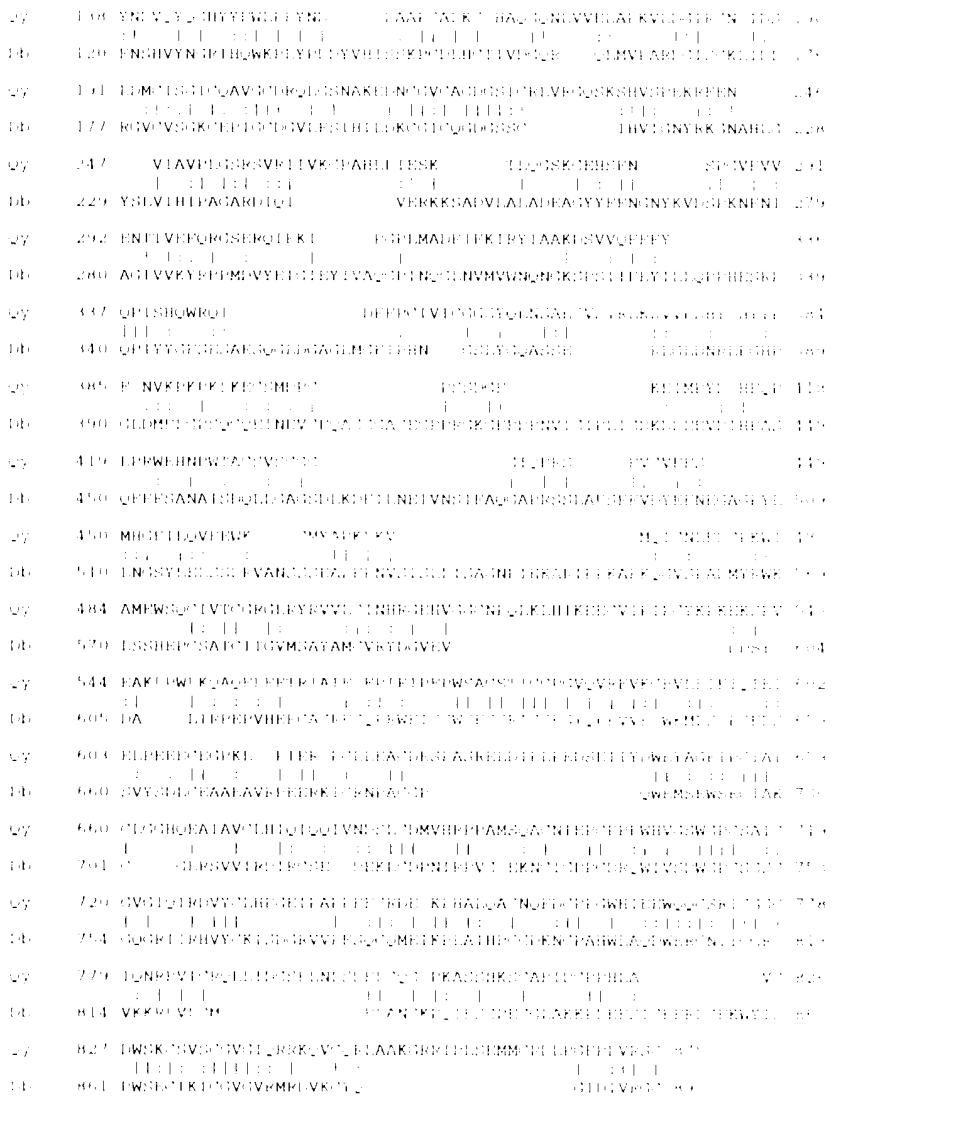
Qy 194 GIGGGGAVGGGFGGVGAGFGGVGAGFGGFGVFGGVGGAVGGGIAVGL 242

RESULT 5
T00260
hypothetical protein KIAA0605 - human
C;Species: Homo sapiens (man)
C;Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 21-Jul-2000
C;Accession: T00260
R;Nakase, T.; Ishikawa, K.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.; Ohara,
DNA Res. 5, 31-39, 1998
A;Title: Prediction of the coding sequences of unidentified human genes. IX. The comp
A;Reference number: Z14086; MUID:98290545
A;Accession: T00260
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-951 <NAG>
A;Cross-references: EMBL:AB011177; NID:g3043733; PIDN:BAA25541.1; PID:g3043734
A;Experimental source: brain
C;Genetics:
A;Note: KIAA0605
C;Superfamily: thrombospondin type 1 repeat homology
F;46-106/Domain: thrombospondin type 1 repeat homology <THR1>

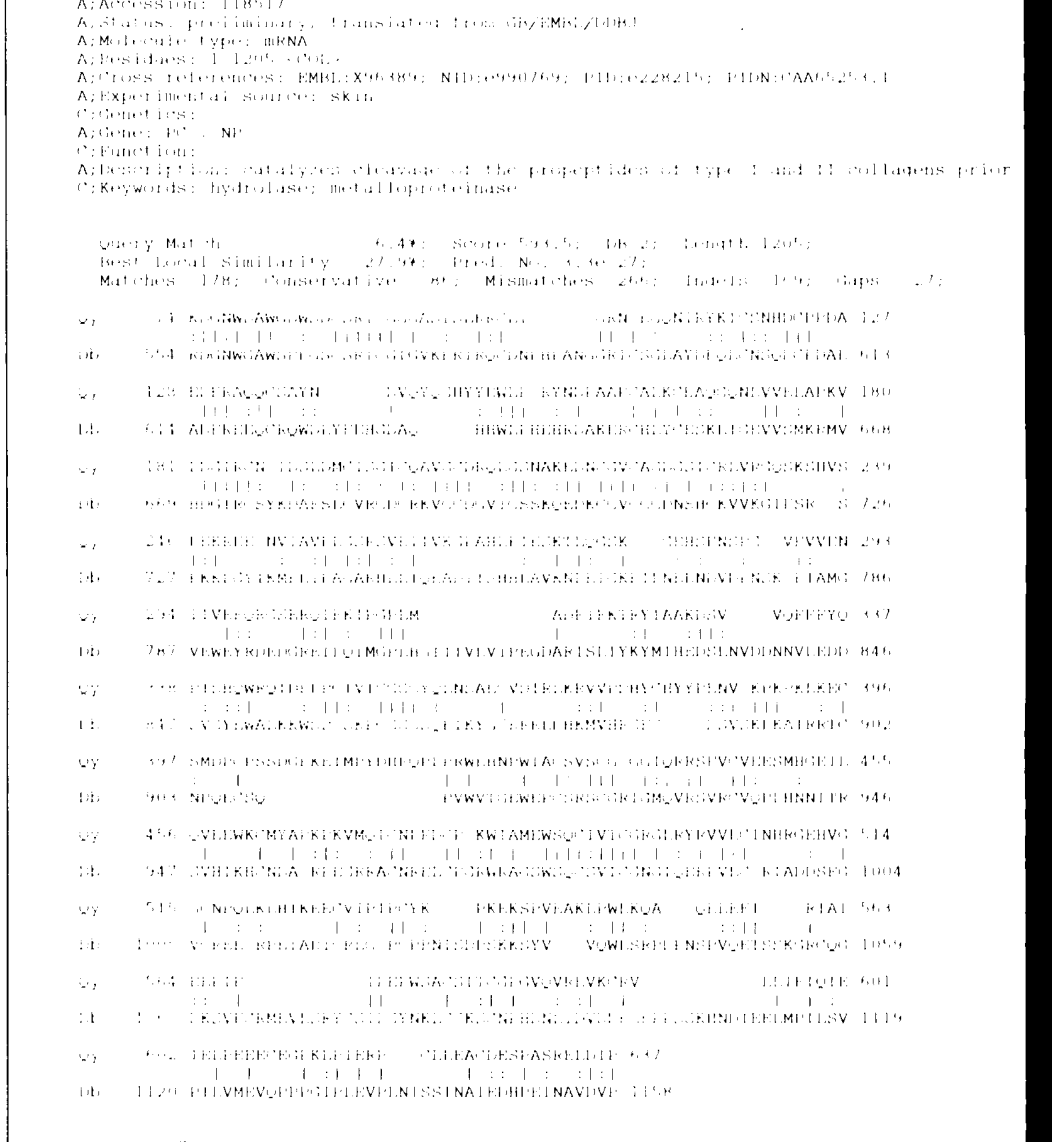
Query Match 7.5%; Score 700; DB 2; Length 951;
Best Local Similarity 24.4%; Pred. No. 1.1e-33;
Matches 214; Conservative 116; Mismatches 351; Indels 294; Gaps 43;

Qy 22 AYFLPFTALFPQGSFLPTTTGEQFLYRYFDQTSRNTIRSDEDSYDGNWDAWGDWSD---G 87
Db 11 AWFLVFAV - VASFIVLSSLNSFNSFSGLLAIAFW WFWFWFWIAF 59

Qy 88 [illegible] 137
Db 60 [illegible] 119



RESULT 6
I18517
procollagen N-endopeptidase (EC 3.4.24.14) I Bovine
N;Alternate names: procollagen N-proteinase
C;Species: Bos primigenius taurus (cattle)
C;Date: 15-Oct-1999 #Sequence revision 15-Oct-1999 #Text change 15-Oct-1999
C;Accession: I18517
R;Colige, A.; Nusgens, B.V.; Lapiere, C.M.
Submitted to the EMBL Data Library, February 1996
A;Description: Cloning of the cDNA of the bovine procollagen I N-proteinase.
A;Reference number: Z18641
A;Accession: I18517
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1205 <Col>
A;Cross-references: EMBL:X96389; NID:g990769; PID:g2282175; PIDN:CAA65253.1
A;Experimental source: skin
C;Genetics:
A;Gene: PCP - NP
C;Function:
A;Description: catalyzes cleavage of the propeptides of type I and II collagens prior
C;Keywords: hydrolase; metalloproteinase



RESULT 7
I14764
hypothetical protein DKFZp434B2041 human (fragment)
C;Species: Homo sapiens (man)
C;Date: 20-Sep-1999 #sequence revision 20-Sep-1999 #text change 20-Sep-1999
C;Accession: I14764
R;Wambutt, R.; Beaubier, I.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
Submitted to the Protein Sequence Database, August 1999
A;Reference number: Z18781
A;Accession: I14764
A;Status: preliminary

```
A;Molecule type: mRNA
A;Residues: 1-898 <WAM>
A;Cross-references: EMBL:AL110226
A;Experimental source: adult testis; clone DKFZp434H204
C;Genetics:
A;Note: DKFZp434H204.1


  Query Match              6.3%;  Score 593;  DB 2;  Length 898;
  Best Local Similarity   24.4%;  Pred. No. 2.4e-27;
  Matches  269;  Conservative  111;  Mismatches  374;  Indels  354;  Gaps   58;

Qy     688 VHRPPAMSQACNIEPCPP---RWHVGSWGPCSAICGVGIQTRDVYCLHPGETPAPPEEC- 743
           :||      :|   :  ||     || | |  |: ||| |:| ::||||      |   | |
Db      20 IRP--   FAGCHPEVPPPVPSWHYGPWTPCTVTCGFGVQPQNVYCLEPQAGPVDEEHCD 74

Qy     744      EDEKPHALQACNQFDCPPCWHIEEWQQCSRTCG GGTQNRRVTCEQLLTDGSFLNL 798
                |::    : |::  ||  |   ||| || :|| ||   | | | : :        |
Db      75 PLGRPDDQQ---PKCSFQPCPARWWAGEWQLCSSSCGPGGLSRRAVLCIPSVCLDPQSAL 131

Qy     799 PDEICQAVEKAGHPSTAR IGTPPDLAVGSWSPCSVSCGVGIQERFLVCLE   IAAFC 853
           |:   |:  :   | |   ||   |||:||:|||:|| | ||| :|    :
Db     132 EPPACEHLPRPPTETPCNRHVPCPATWAVGNWSQCSVTCGEGTQRRNVLCINDTGVPCDE 191

Qy     854 RRIPLSEMMCRDLPGFPLVRSCQMPFCSKTKSSMKIEL..SQCPQGISVQPVYTQPEEPP 913
           : | ||: |      ||   |: |              | ||:
Db     192 AQQPAGEVTC----SLPL   CRWP           LSILSFL                  216

Qy     914 INLTTGSRAYLLFNISVIIL.GVEELPPLQAIFS S TGNSFPL  IIG-GSLKTHSLA 972
             || :: | | :  :                                        | ||
Db     217  QPPSPPCPHFLSNFAFFP                                    HHLA 270

Qy     973 AGGGVYSYTIAT'A_ITVVIFLI'IINRLIAFLAICFEIMSLIPMLHSEANFLGVIWHEM 1032
             |    .: :     ||   |  |  : ||    :
Db     240 P------RPSPASSPKP   ----GTMGNAIEEEA---PELDLPCPVPVD--------   273

Qy    1033 RQMWNNENGTYIIPKHIPQLTPLALLDHTSNCAGTPNLWELEGKLPAAVELGAYLMEI 1092
             | | : ::          |  | | :   :|
Db     274 ---------DFYYDYNFIN---------FHEDLSYGPSEEPDL-- ------ ---------- 298

Qy    1093 AQPPEITNNMSLLMEIREVSPLLAFQLITVLVAILAFAAPPTHMQ       WEGIQEEIPP 1147
                    |  ||:                    :  | | :          :     ||
Db     299 ---------  DLAGTGD                KIPPPRSRPAAPSTGSPVPAIEPP 329

Qy    1148 AAQLRCETCSVSQSSQAFNSCPLIEPPGPVLMPQVPFSIRNKIINSPGGNIVYITKE 1207
           ||: |  | | | |     ::           |  | |
Db     330 AAKEEGVLGPWSPSPWPSQAG-  -- -- --RSPPPPSEQ--- - --       TPG 362

Qy    1208 IEVINILCDLIIPSEA------IYIWIK-  DG--------------------TLLQPSV 1238
             :|| | :  |:  |       : :| :   ||                    : | |
Db     363 NPLINFLPEEDTPIGAPDLGLPSLSWPFRVSTDGLQTPAIPESQNLPPVGKDSQSQLPPPW 422

Qy    1239 EIILPIPGPEQTQNPIPPEEQPIYPCSVANHLGSIVESSVLYADAPVILSVDENITKPEH 1298
           :  | | ::   :  | :|      | ||      |  |    | :  |    : |
Db     423 R---DRINEVEKGDEEPPKDRG -  - APHL--PPRPSSIL-  - PPLSPVGSIHSSPSP 467

Qy    1299 NHLSVVVPGIV       PAALPANVFIPGFVPGIV   -       LPNITWLKEGGGLS 1352
           :   || |     |  ||    |  ||                  | ||   |  ||
Db     468 DVAELWTGGIVAWEPALEGGLGPVDSELWPIVGVASLLPPPIAPLPEMK  VRDSSLEPG 525

Qy    1353 NVGLLP  NGSLLLQNVSLENEGIYVCIATNAIGKAVATSVLHLLE-------RRWPESR 1403
             |      ||  || :::   :|::      ||        |: :       :  |||
Db     526 IPSPPAPCPGGWDQIVAV  WIPLPTTLIGLG       - HMPEPALNPCPKGQPESL 575

Qy    1404 IVPLQGHKKYI DQAINTEKTNSNP PTGEP          PPQEPP      WPEGNWSHC 1456
           :   : :  | :: |:  | : :|       || :|      |: |||| |
Db     576 SPEVPLSSSRLLSIPAWDSPANSHEVPETQPLAPSLAEAGPPADPIVVKNAGWQAGNWSHC 635

Qy    1457 SATCGHGGARIQRR-QVVMANGLEVSEALCDHIQKPLAGFEP---CNIRDCPARWFTSVW 1512
           | ||| ||| : || :|    ::      |    :|    :|  :|:| | | :  |
Db     636 SITCG-LGA VWEPVPCSSGRDED-----CAPAGKP -- QPARRCHLRPC AIWHSGNW 683

Qy    1464 SQCSVSCGEGYHSPQVICKPIEANCIVQVVSPPACAD-KDRPILCRKPCPGHWCVQWEPGN 1522
```

```
           |:|| ||| |   | | |  |:    :: : |  | |    :|   :||   ||: |   :
Db     684 SKCSESCGGGSSVRDVQCVDIR---DLSPLRPPHCQPCPAKPPAPRPPCGAQPCLSWYTSS 740

Qy    1523  -PCPCPCMSEAVRMQQSETACQDINSCDSNCDPERKPTLPENCTCGACDVCWHTGTWRPC 1581
              |  | |     ||| |          |:: ||  | | :  |   ||| |
Db     741 WREESEAGCGGH-- QQKDLVTC---PEPGLCPEEALKPNTIKPCNTHPC TQWVVGPWGQC 793

Qy    1582 TAACGRGPQSRPVVCIQTPSCPPVA--PPPCVQPPPDISWRHC---------LGPSCDRDCT 1630
             :| || | | | | |::|::  |      |  :  | | | |              | |:||
Db     794 SAPCGGGVQKRLVECVNTQTGLPPEECPQTGHEAWEEGSRPCGTECCPPVEPPPCPEDRL 853

Qy    1631 DITHYCMFVKHLNLCSLDRYKQRCCQSC 1688
             :  :|  :: |  | |   : |||:||
Db     854 -SFGPCEPILRLLGRCQLPPIRTQCCRSC 880


RESULT   8
T47158
hypothetical protein DKFZp762C1110.1   human (fragment)
C;Species: Homo sapiens (man)
[illegible] 20-Apr-[illegible] #sequence_revision 20-Apr-[illegible] #text_change 20-Apr-2000
C;Accession: T47158
R;Blum, H.; Bauersachs, S.; Mewes, H.W.; Weil, B.; Wiemann, S.
submitted to the Protein Sequence Database, March 2000
A;Reference number: S24329
A;Accession: T47158
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-550 <AAA>
A;Cross-references: EMBL:AL162080
A;Experimental source: adult melanocyte (NHEM-A II line); clone DKFZp762C1110
C;Genetics:
A;Note: DKFZp762C1110.1


  Query Match              6.1%;  Score 571.5;  DB 2;  Length 550;
  Best Local Similarity   [illegible];  Pred. No. 2.5e-26;
  Matches  136;  Conservative   77;  Mismatches  164;  Indels  105;  Gaps   19;

Qy      76 GNWDAWGDWSDCSRTCGGGASVSLRRCLI------GRNCEGQNIRYKTCSNHDCPP-DAE 128
             |:|  || | ||||||||| |:|| |         |: ||| :|||:|:  ||| : :
Db     143 GSWGMWGCWGLCSRSTRTRSVVLQMSLTNIVEENGSCTGGREVKYRACSNLDCPIGNNSK 202

Qy     129 DFRAQQFSAYNCVQY -- QGRYYFMWIPRYN  DFAAAPCALECHAQGQNEVVELAFEVLD 186
             || :|| |:|:          |  ||:|:|    |  | | | |:|     | |||:|
Db     203 TPREEQCTPAPNFDCPAPDRCPAVFWIPFYATVCPPEGCKIQQAPGCIPYFTVLQPEVVD 262

Qy     187 CIPCNTDSTLPMCISGTGAVGCDRQLGCNAEEQNCCVCACGEGSPCPLVPGGPKSHVCPPK 246
             || |: ||  :|: | |  |||| : | |  | |:|| ||||||: : |   | | :
Db     263 CIPCSPDSTSVCVQGQGVKAGCDRCIPSPEEPPPGCVQGGNGSPCPKKISG---SVISAKP 319

Qy     247 PEENVIAVPLGSPWVPITVKGPAHLPTDSKTIQGGREGHESP    NSPSVPVVEN--  SI 304
             :|| :| |: ::          :: :  :||:   ||     : | :::       :
Db     320 GYHDTTTPTGGAINIL          VKQSNQKGSNNNGSPLAAKAAGSGYLGNGDYTLS 369

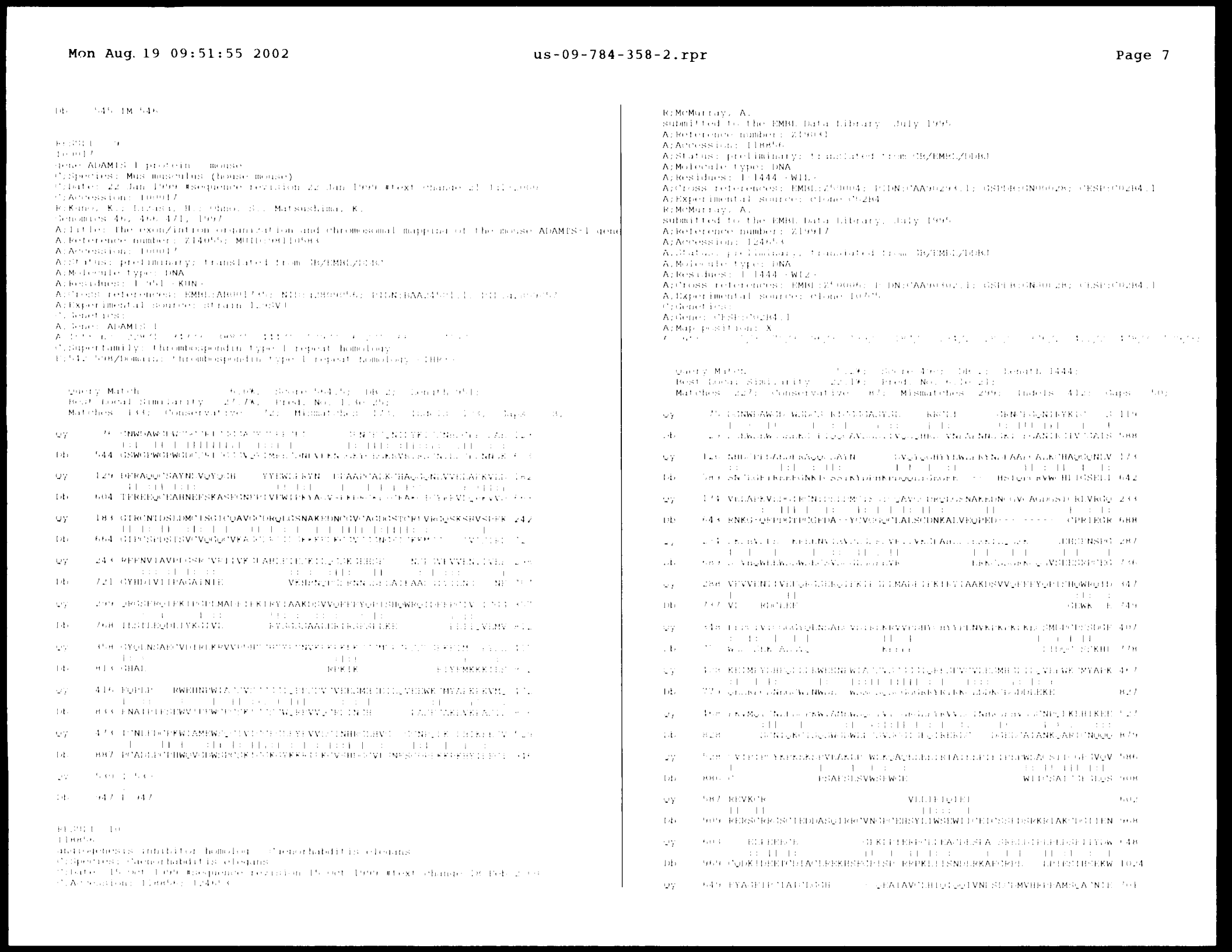
Qy     305 TVEPQPGSSEMQTPKIPGPLMAGPSPE- -IRGIAAPDSVVQEEYQPTSHQWRQTDFFPC 363
             |:|                    | :::    ||: :  :: :   : |:          |
Db     370 TLP                  QDIMYKGVVLEYSGSSAALEKIRSFSPLKE- -- -- PL 404

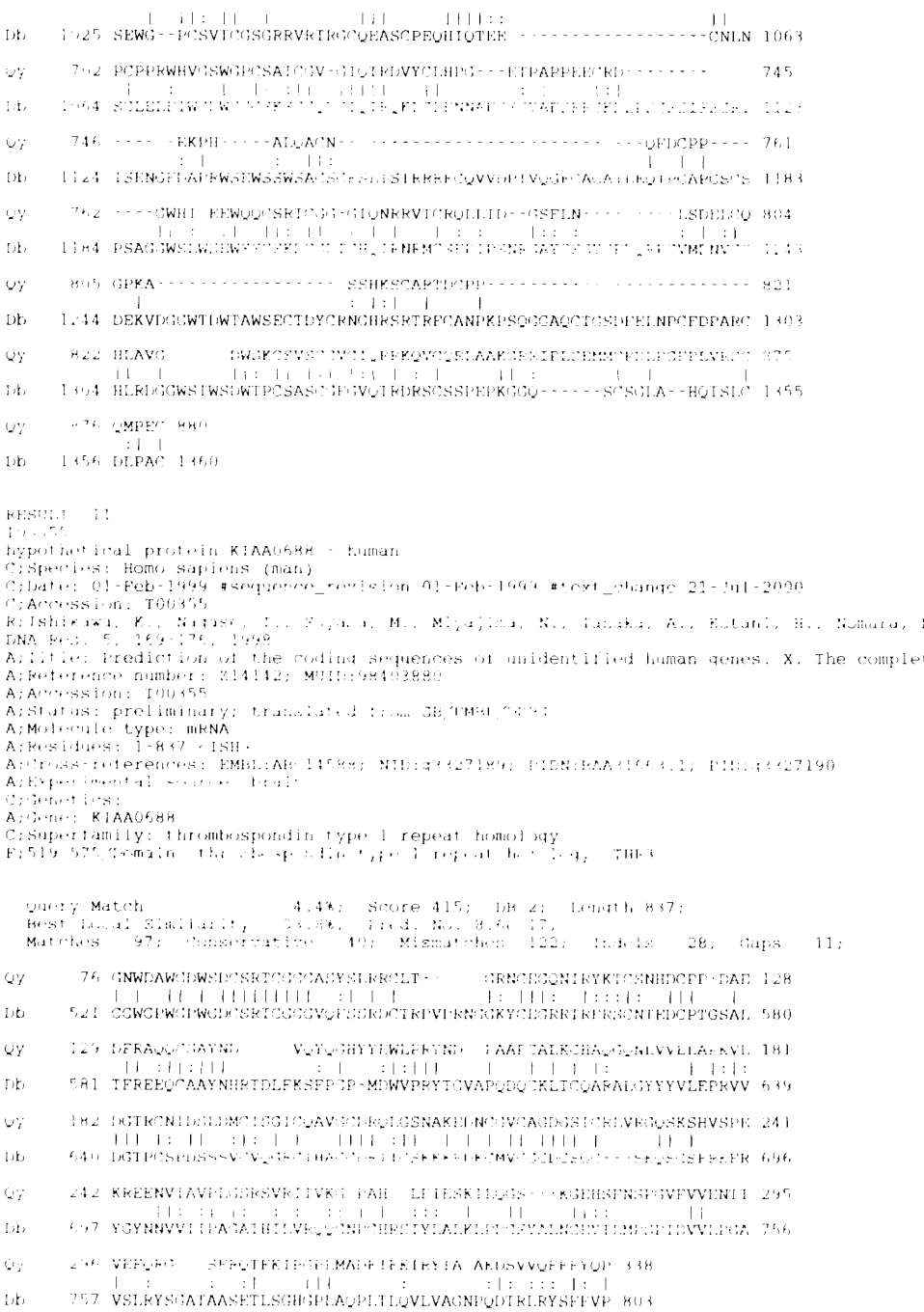
Qy     364 IVIGGGYQLNSAECVDIRLKKVVSTDHTLEYPENVKPKPPKLKKCSMDPCPSSGPRKEIM 423
             |:       |    : :::      | ::          [ |: | :  |:
Db     405 TIQV ----LEVGNALKLKLE       YSTLV         KKKKKEGLNAIPT-  ----- 437

Qy     424 FYDHPQPLPEWEHNEWTACSVSCGGGIQERSVVCVPPSMPGPILQVPPWKCSKVADPPPKVM 483
                  |    |  || :|  | :||   |    ::|:          :|    ||
Db     438         FSAWVIEEWGACSESCEGGWQKRLVEC  KDINEGLAS    ECAKEVEPAST 484

Qy     484 QTCNLPGCPEWIAWEWSACSVTCGEGIKYRVVCVPPSMPGPILQVPPWKCSKVADPPPKVM 543
             : |    ||:|   ||| |: |||:| : | : |:: | |  :    |:|  |  | : |
Db     485 RPCADHPCPQWQLGEWSSCSKSCGKGFKFRSLECLSHDGCVELSHESKCPPLKEPKHPIDPC 544

Qy     544 VI 550
             :
```



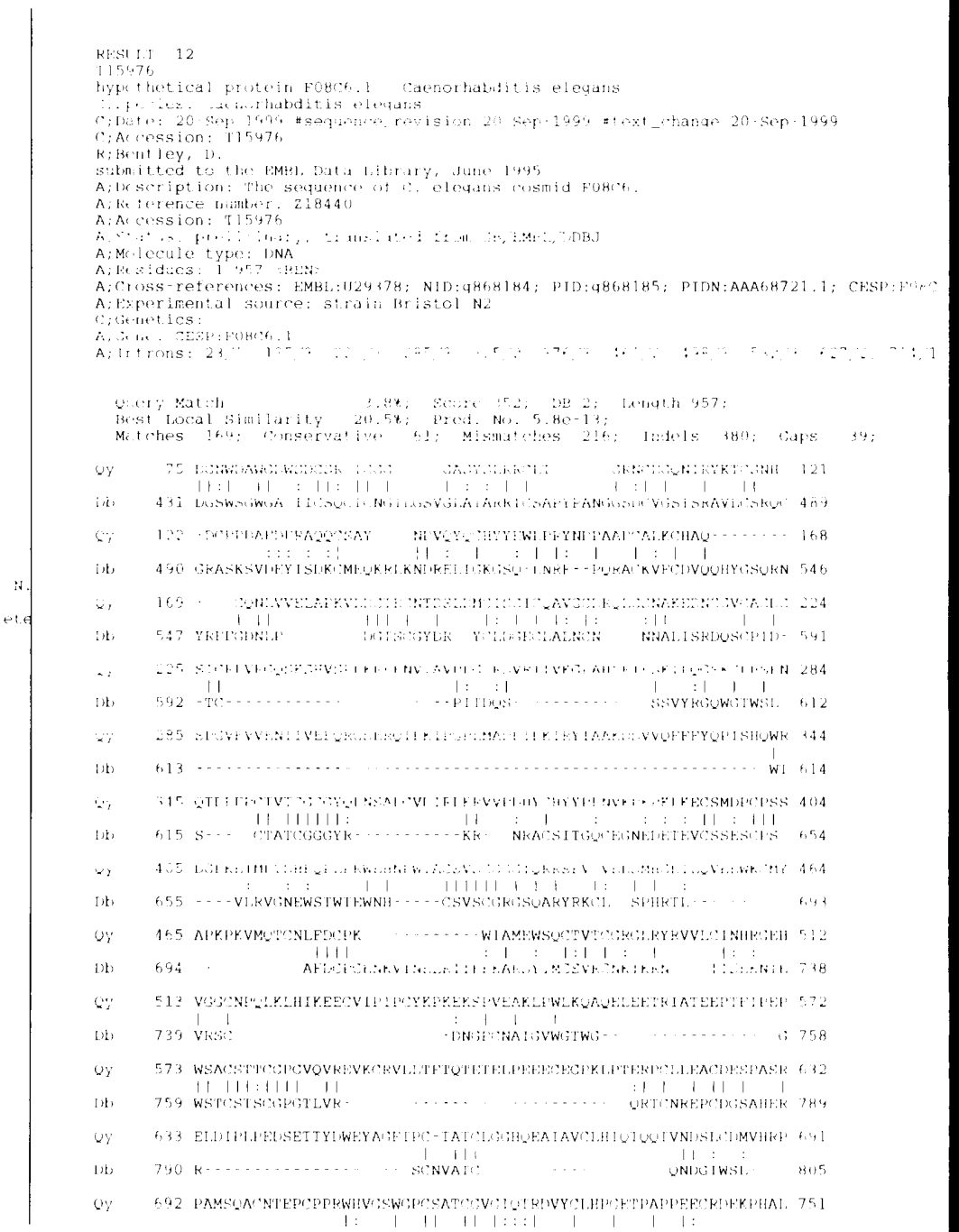


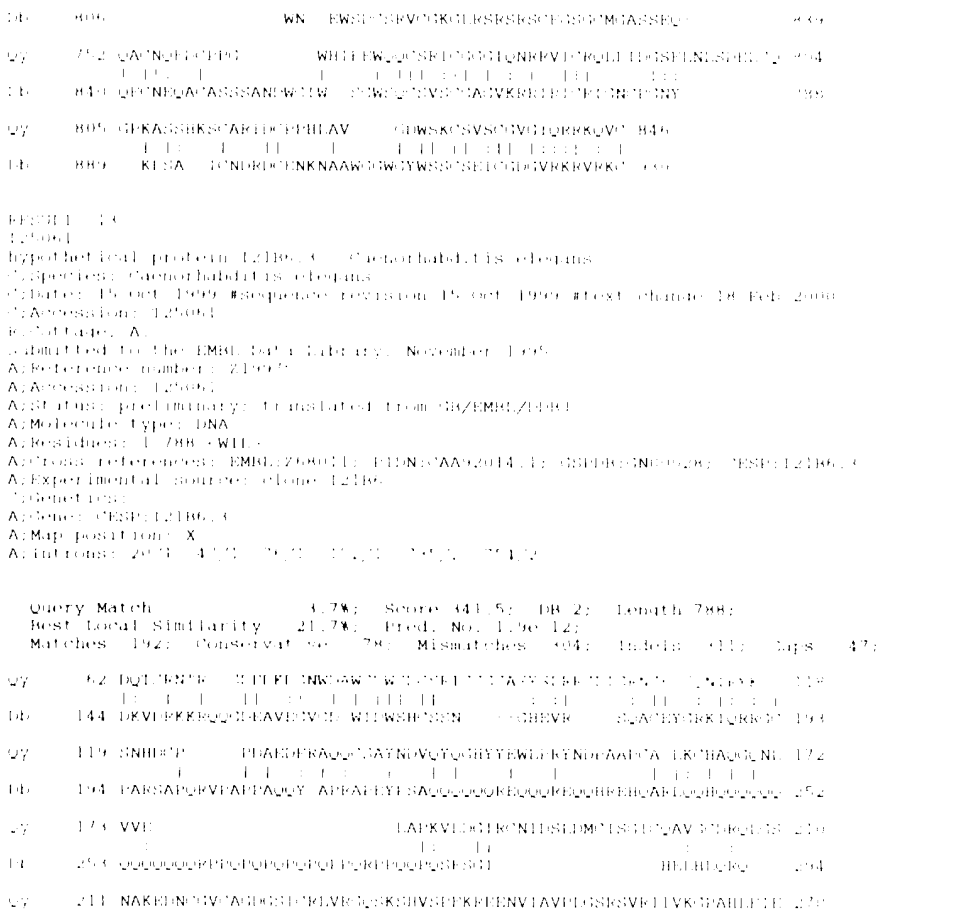
RESULT 11
I15976
hypothetical protein KIAA0688 - human
C;Species: Homo sapiens (man)
C;Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 21-Jul-2000
C;Accession: T00355
R;Ishikawa, K., Nagase, T., Fujiwara, M., Miyajima, N., Tanaka, A., Kotani, H., Nomura, N.
DNA Res. 5, 169-176, 1998
A;Title: Prediction of the coding sequences of unidentified human genes. X. The complete
A;Reference number: Z14142; MUID:98403880
A;Accession: T00355
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-837 <ISH>
A;Cross-references: EMBL:AB011549; NID:g7427190; PIDN:BAA31944.1; PID:g7427190
A;Experimental source: brain
C;Genetics:
A;Gene: KIAA0688
C;Superfamily: thrombospondin type 1 repeat homology
F;519-575/Domain: thrombospondin type 1 repeat homology; THBS

Query Match 4.4%; Score 415; DB 2; Length 837;
Best Local Similarity 23.8%; Pred. No. 8.6e-17;
Matches 97; Conservative 40; Mismatches 122; Indels 28; Gaps 11;

RESULT 12
I15976
hypothetical protein F08C6.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C;Accession: I15976
R;Bentley, D.
submitted to the EMBL Data Library, June 1995
A;Description: The sequence of C. elegans cosmid F08C6.
A;Reference number: Z18440
A;Accession: I15976
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-957 <BEN>
A;Cross-references: EMBL:U29478; NID:g868184; PID:g868185; PIDN:AAA68721.1; CESP:F08C
A;Experimental source: strain Bristol N2
C;Genetics:
A;Gene: CESP:F08C6.1
A;Introns: 23/3 107/3 221/3 285/3 475/3 576/3 618/3 648/3 742/3 787/1

Query Match 2.0%; Score 192; DB 2; Length 957;
Best Local Similarity 20.5%; Pred. No. 5.8e-13;
Matches 169; Conservative 61; Mismatches 216; Indels 80; Gaps 39;



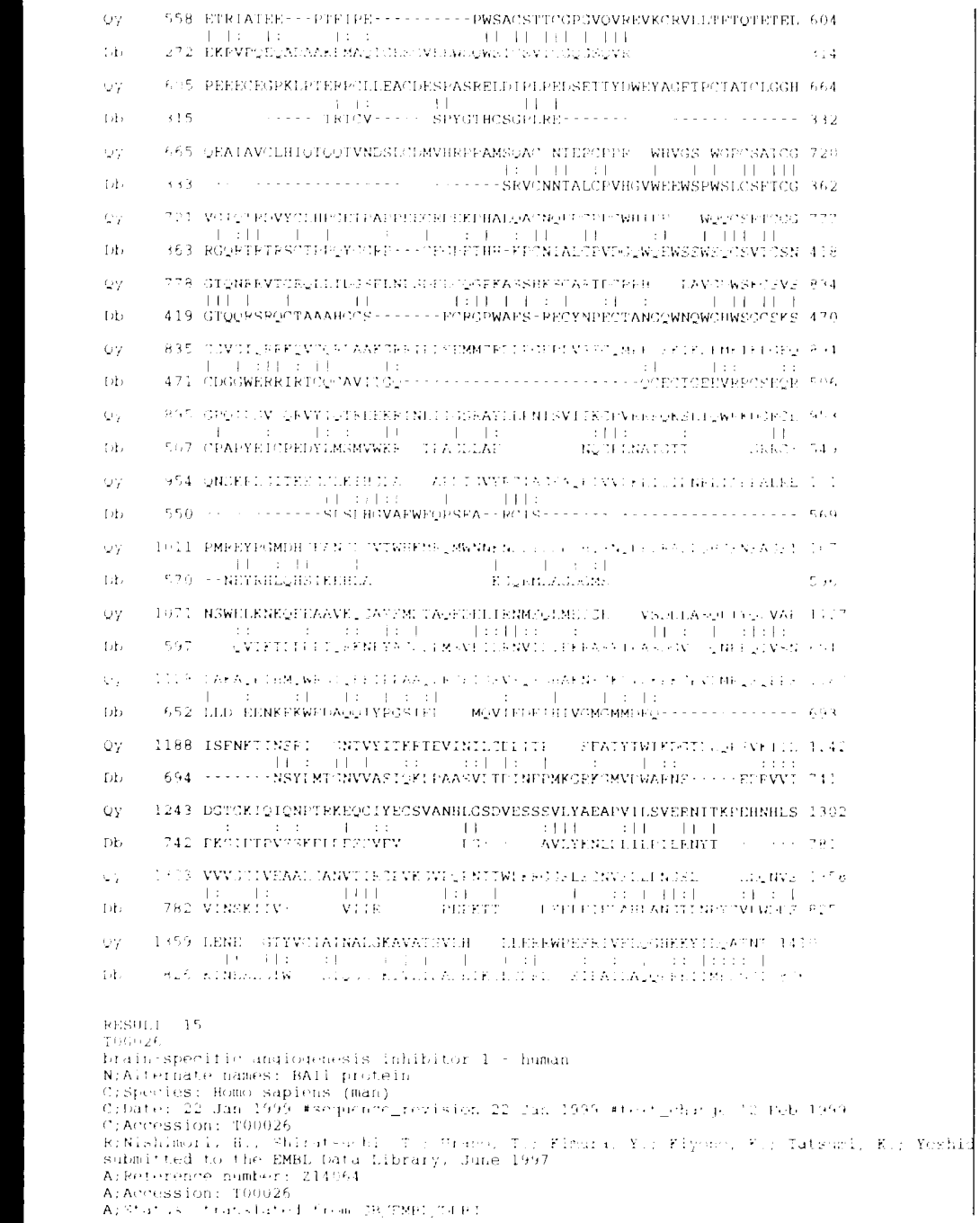


RESULT 13
I25061
hypothetical protein I25061 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #Sequence revision 15-Oct-1999 #text change 28-Feb-2000
C;Accession: I25061
R;Cottage, A.
Submitted to the EMBL Data Library, November 1995
A;Reference number: Z19977
A;Accession: I25061
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-788 <WIL>
A;Cross-references: EMBL:Z68011; PIDN:CAA92014.1; GSPDB:GN00206; ESP:L21861.3
A;Experimental source: clone L2186
C;Genetics:
A;Gene: CESP:L2186.3
A;Map position: X
A;Introns: 20/3; 41/2; 79/2; 121/3; 247/2; 271/2

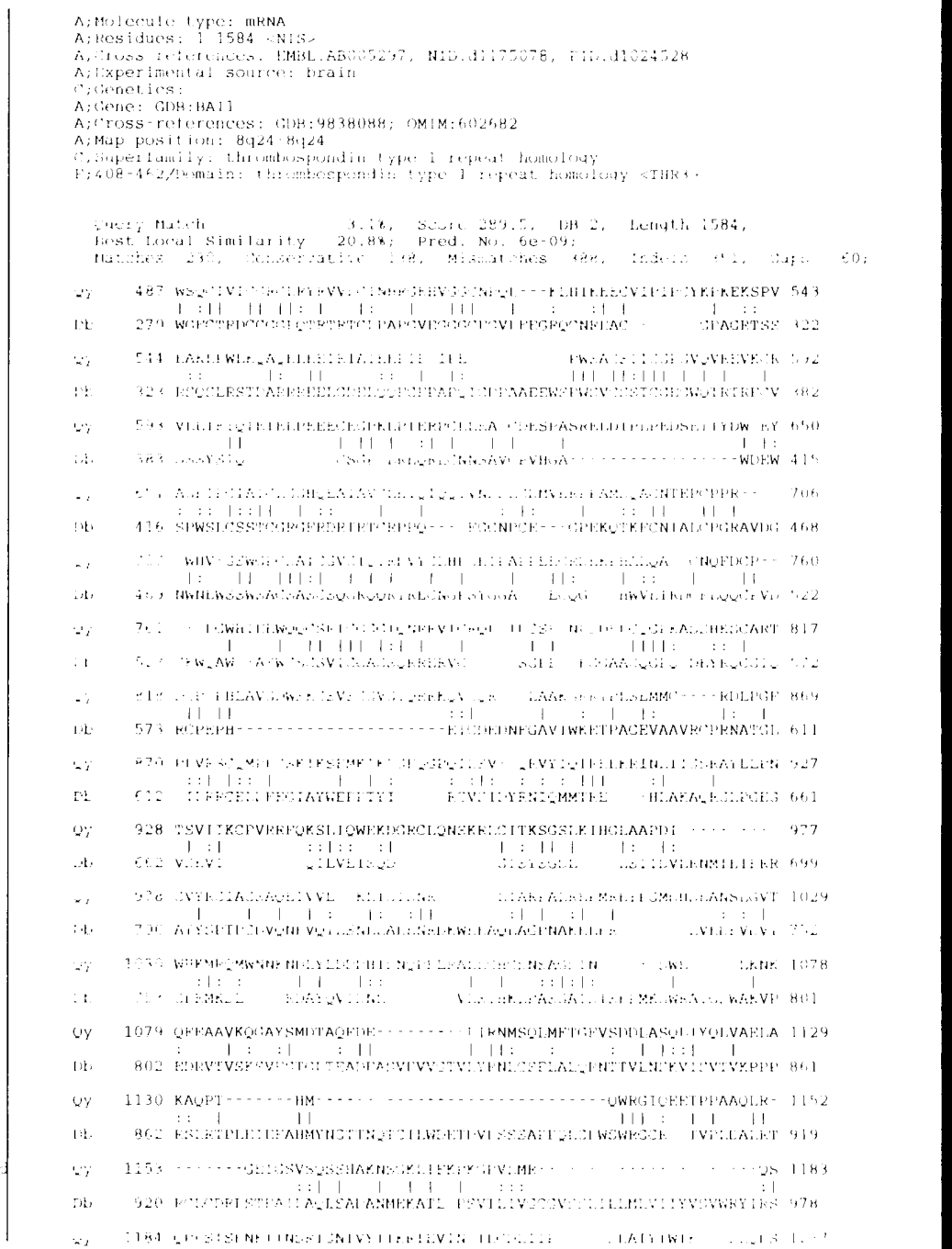
Query Match 4.7%; Score 341.5; DB 2; Length 788;
Best Local Similarity 21.7%; Pred. No. 1.9e-12;
Matches 192; Conservative 98; Mismatches 304; Indels 111; Gaps 47;

RESULT 14
I00028
brain-specific angiogenesis inhibitor 1 - human
N;Alternate names: BAI1 protein
C;Species: Homo sapiens (man)
C;Date: 22-Jan-1999 #sequence revision 22-Jan-1999 #text change 21-Jul-2000
C;Accession: I00028
R;Shiratsuchi, T.; Nishimori, H.; Ichise, H.; Nakamura, Y.; Tokino, T.
Cytogenet. Cell Genet. 79, 103-108, 1997
A;Title: Cloning and characterization of BAI2 and BAI3, novel genes homologous to bra
A;Reference number: Z14566; MUID:98194217
A;Accession: I00028
A;Status: translated; 1 sequence; DBSOURCE_DDBJ
A;Molecule type: mRNA
A;Residues: 1-1522 <SHI>
A;Cross-references: EMBL:AB000209; NID:g3021709; PIDN:BAA25363.1; PID:g4021701
A;Experimental source: brain
C;Genetics:
A;Gene: GDB:BAI3
A;Cross-references: GDB:983809; MIM:602684
A;Map position: 6q12-6q12
C;Superfamily: thrombospondin type 1 repeat homology
F;344-398/Domain: thrombospondin type 1 repeat homology <TSR3>

Query Match 4.1%; Score 292.5; DB 2; Length 1522;
Best Local Similarity 16.5%; Pred. No. 8.8e-09;
Matches 233; Conservative 94; Mismatches 443; Indels 414; Gaps 62;

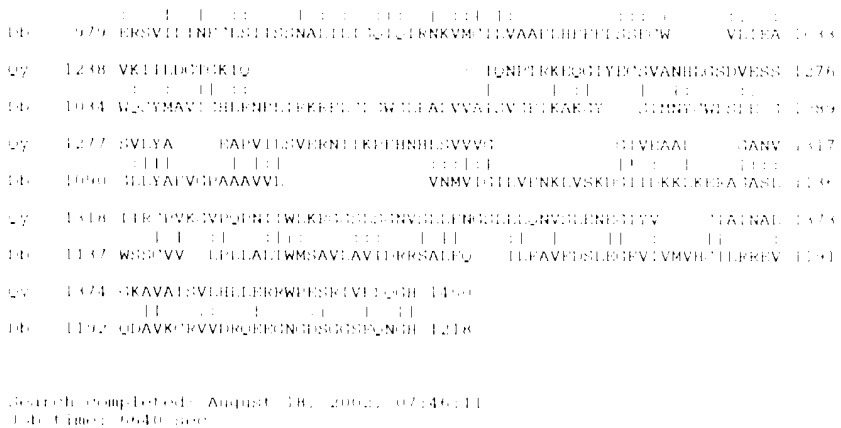


RESULT 15
T00026
brain-specific angiogenesis inhibitor 1 - human
N;Alternate names: BAI1 protein
C;Species: Homo sapiens (man)
C;Date: 22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change 12-Feb-1999
C;Accession: T00026
R;Nishimori, H., Shiratsuchi, T., Urano, T., Kikuno, R., Yamaoka, Y., Flybase, F., Tatsumi, K., Yoshi
submitted to the EMBL Data Library, June 1997
A;Reference number: Z14064
A;Accession: T00026
A;Status: translated from 3D/EMBL/GB



A;Molecule type: mRNA
A;Residues: 1-1584 <NIS>
A;Cross-references: EMBL:AB005297, NID:d1175078, PID:d1024528
A;Experimental source: brain
C;Genetics:
A;Gene: GDB:BAI1
A;Cross-references: GDB:9838088; OMIM:602682
A;Map position: 8q24-8q24
C;Superfamily: thrombospondin type 1 repeat homology
F;408-462/Domain: thrombospondin type 1 repeat homology <THR3>

Query Match 3.1%; Score 289.5; DB 2; Length 1584;
Best Local Similarity 20.8%; Pred. No. 6e-09;
Matches 230; Conservative 198; Mismatches 488; Indels 451; Gaps 50;



Search Completed: August 19, 2002, 07:46:11
Job Time: 6040 secs

GeneCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

SW protein - protein search, using sw model

Run on: August 18, 2002, 05:45:26 ; Search time 25.71 Seconds
(without alignments)
1000.522 Million cell updates/sec

Title: US-09-784-358-2
Perfect score: 9446
Sequence: 1 MASWTSPWWVEINVEMRSP.....ENGTSIDFYK_F^%S^_F^ 441

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 230698 seqs, 241255704 residues

Total number of hits satisfying chosen parameters: 230698

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/2/iaa/*A_*MB_pep:*
2: /cgn2_6/ptodata/2/iaa/*B_*MB_pep:*
3: /cgn2_6/ptodata/2/iaa/*A_*MB_pep:*
4: /cgn2_6/ptodata/2/iaa/*B_*MB_pep:*
5: /cgn2_6/ptodata/2/iaa/backfiles1_pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles2_pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

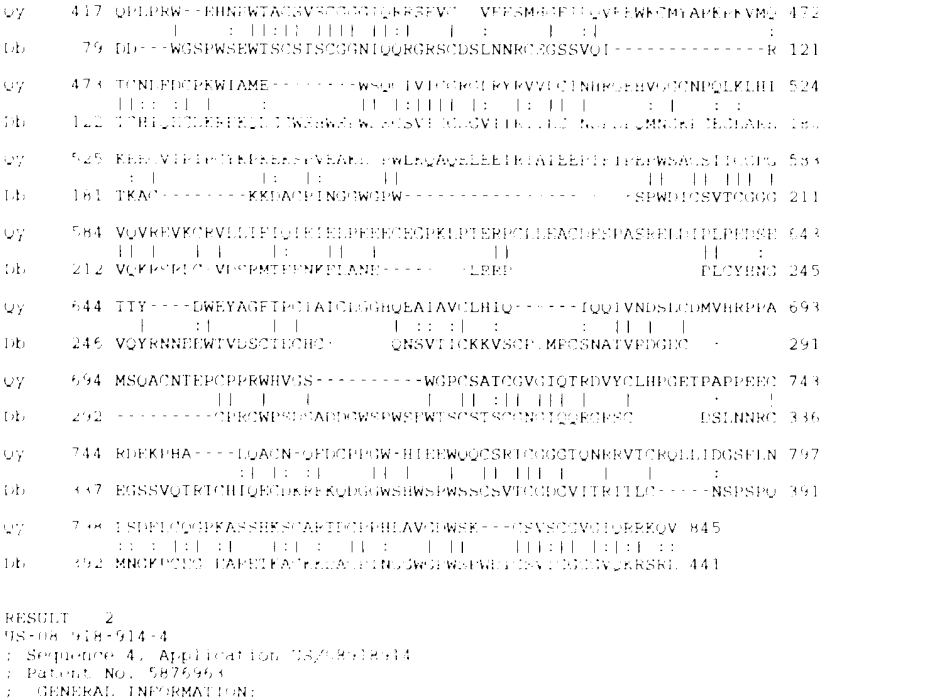
Result No.	Score	Query Match	Length	DB	ID	Description
1	865.5	9.2	441	4	US-08-985-526-4	Sequence 4, Appli
2	841.5	8.7	708	2	US-09-918-914-4	Sequence 4, Appli
3	273	2.9	459	1	US-08-813-288B-15	Sequence 15, Appl
4	265	2.8	807	1	US-07-862-021B-10	Sequence 10, Appl
5	265	2.8	807	1	US-08-813-288B-10	Sequence 10, Appl
6	265	2.8	807	5	PCT-US93-03164-10	Sequence 10, Appl
7	264	2.8	568	1	US-07-862-021A-14	Sequence 14, Appl
8	264	2.8	568	5	PCT-US93-03164-14	Sequence 14, Appl
9	256	2.7	802	1	US-07-862-021B-12	Sequence 12, Appl
10	256	2.7	802	1	US-08-813-288B-12	Sequence 12, Appl
11	256	2.7	802	5	PCT-US93-03164-12	Sequence 12, Appl
12	240	2.5	1179	1	US-08-813-288B-20	Sequence 20, Appl
13	240	2.5	1182	1	US-08-813-288B-19	Sequence 19, Appl
14	213.5	2.3	630	2	US-08-752-807B-14	Sequence 14, Appl
15	207	2.2	607	2	US-08-752-807B-12	Sequence 12, Appl
16	203	2.2	805	4	US-08-985-526-44	Sequence 44, Appl
17	202	2.2	805	1	US-08-433-861-5	Sequence 5, Appli
18	202	2.2	805	4	US-08-103-829B-5	Sequence 5, Appli
19	202	2.2	1367	1	US-07-814-593-4	Sequence 4, Appli
20	202	2.2	1367	1	US-07-977-451-6	Sequence 6, Appli
21	202	2.2	1367	1	US-07-946-507-4	Sequence 4, Appli
22	202	2.2	1367	1	US-08-252-517-6	Sequence 6, Appli
23	202	2.2	1367	1	US-07-906-497A-6	Sequence 6, Appli
24	202	2.2	1367	1	US-08-601-891-6	Sequence 6, Appli
25	202	2.2	1367	2	US-08-443-861-2	Sequence 2, Appli
26	202	2.2	1367	2	US-09-021-824-6	Sequence 6, Appli
27	202	2.2	1367	4	US-08-103-829B-2	Sequence 2, Appli
28	202	2.2	1367	5	PCT-US92-02750-8	Sequence 8, Appli
29	202	2.2	1367	5	PCT-US92-05401-6	Sequence 6, Appli
30	202	2.2	1367	5	PCT-US92-09893-6	Sequence 6, Appli
31	202	2.2	4544	1	US-08-469-486-52	Sequence 52, Appl
32	202	2.2	4544	2	US-08-469-658-52	Sequence 52, Appl
33	200.5	2.1	768	2	US-08-874-678-1	Sequence 1, Appli
34	200.5	2.1	768	3	US-08-643-839-1	Sequence 1, Appli
35	200.5	2.1	768	4	US-09-051-383-24	Sequence 24, Appl
36	200.5	2.1	780	1	US-08-232-538-14	Sequence 14, Appl
37	200.5	2.1	786	2	US-08-286-164-14	Sequence 14, Appl
38	200.5	2.1	1091	3	US-08-986-485-5	Sequence 5, Appli
39	200.5	2.1	1297	4	US-09-540-245A-17	Sequence 17, Appl
40	200.5	2.1	1338	3	US-08-750-141A-3	Sequence 3, Appli
41	200	2.1	239	5	PCT-US93-01052-1	Sequence 1, Appli
42	198.5	2.1	612	2	US-08-752-807B-11	Sequence 11, Appl
43	198.5	2.1	1268	4	US-08-506-299B-28	Sequence 28, Appl
44	196	2.1	2476	2	US-08-276-967-2	Sequence 2, Appli
45	196	2.1	1411	1	US-08-340-011-5	Sequence 5, Appli

ALIGNMENTS

RESULT 1
US-08-985-526-4
; Sequence 4, Application US/08985526
; Patent No. 6080728
; GENERAL INFORMATION:
; APPLICANT: Mixson, James A
; TITLE OF INVENTION: CARRIER RNA COMPLEXES CONTAINING DNA
; TITLE OF INVENTION: THERAPY
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Connolly, Bove, Lodge, & Hutz
; STREET: 1220 Market Street, P.O. Box 2207
; CITY: Wilmington
; STATE: Delaware
; COUNTRY: U.S.A.
; ZIP: 19899
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/985,526
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/108,845
; FILING DATE: 16-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: McMorrow Jr., Robert G
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (302) 658-9141
; TELEFAX: (302) 658-5614
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 441 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-985-526-4

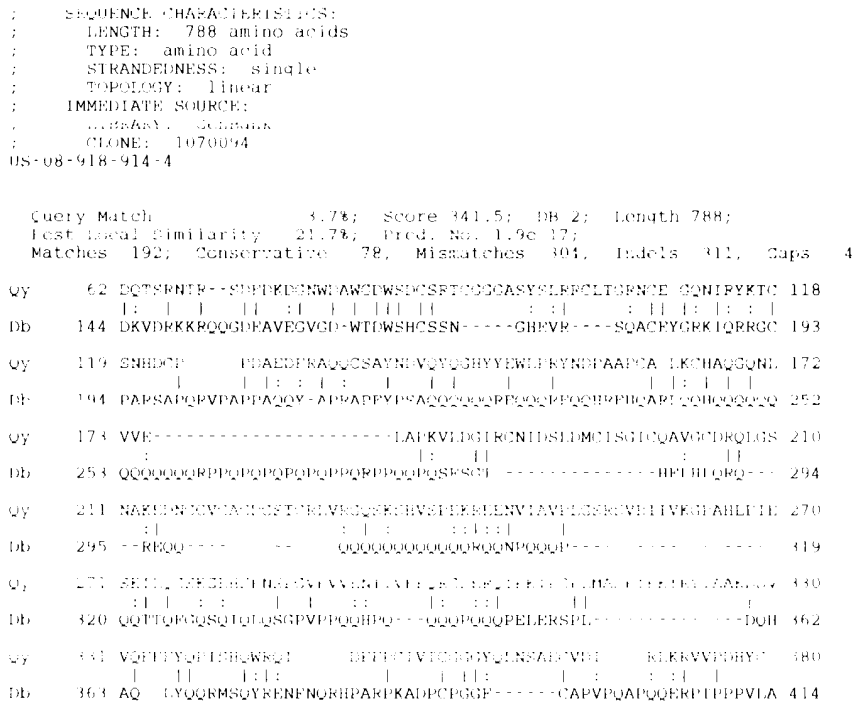
Query Match 9.2%; Score 865.5; DB 4; Length 441;
Best Local Similarity 24.9%; Pred. No. 1.1e-19;
Matches 120; Conservative 71; Mismatches 100; Indels 69; Gaps 27;

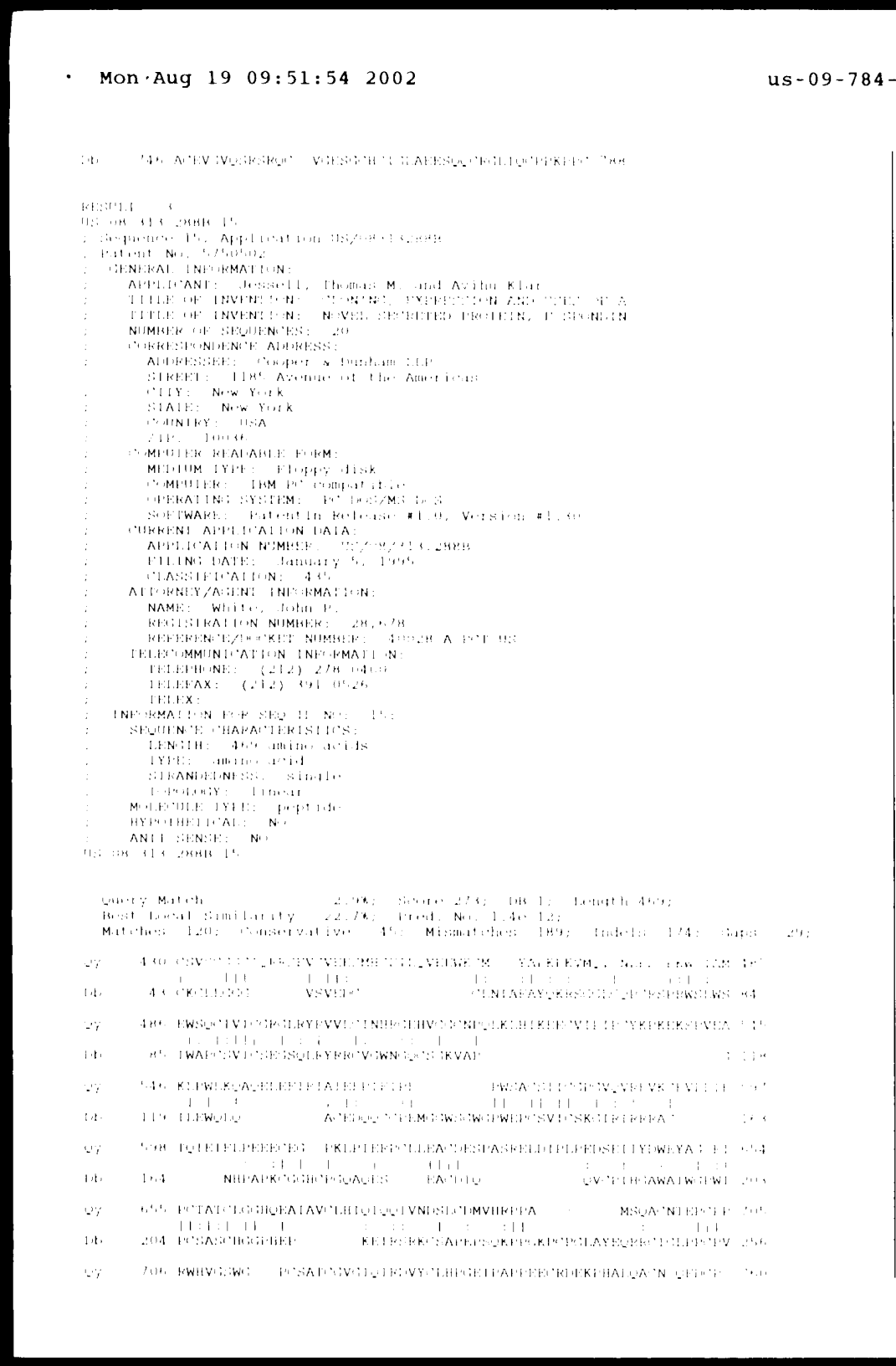
Qy 1 MASWTSPWWVEINVEMRSPDFDNFLNNYSLYDRSLIENSPVRETVDDF 416
Db 27 NNEEWIIDVSCIE CHC QNSVIICKKVSCFIMHCSNAIVHHCECCLSCWHSDSA 78



RESULT 2
US-08-918-914-4
; Sequence 4, Application US/08/918,914
; Patent No. 5876963
; GENERAL INFORMATION:
; APPLICANT: Mitchell, Peter
; APPLICANT: Hutchinson, Nancy
; APPLICANT: Lawton, Michael
; APPLICANT: Magna, Holly
; APPLICANT: Yocum, Sue
; APPLICANT: Murry, Lynn E.
; TITLE OF INVENTION: HUMAN NUCLEOTIDE PYROPHOSPHORYLASE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 4174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/918,914
; FILING DATE: Filed Herewith
; PRIOR APPLICATION DATA:
; APPLICANT:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0364
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415 845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 788 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: Stomach
; CLONE: 1070094
US-08-918-914-4

Query Match 3.7%; Score 341.5; DB 2; Length 788;
Best Local Similarity 21.7%; Prob. No. 1.9e-17;
Matches 192; Conservative 78; Mismatches 301; Indels 311; Gaps 42;





RESULT 3
US/08/413,208B/1%
; Sequence 1%, Application US/08413208B
; Patent No. 5750502
; GENERAL INFORMATION:
; APPLICANT: Jessell, Thomas M. and Avihu Klar
; TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A
; TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC DOS/MS DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/413,208B
; FILING DATE: January 5, 1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 40928 A PCT US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278 0400
; TELEFAX: (212) 391 0525
; TELEX:
; INFORMATION FOR SEQ ID NO: 1%:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 400 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US/08/413,208B/1%

Query Match 2.7%; Score 27; DB 1; Length 400;
Best Local Similarity 22.7%; Pred. No. 1.4e+12;
Matches 120; Conservative 45; Mismatches 189; Indels 174; Gaps 29;

Qy 430 CSVSSSSSLRSLEVTVELSMSTSLVELGE-M EALELEVML- Seq. raw 1025 to
Db 45 CPCLLDSG VSVEPP CLNLAEAYQRRSQDLGLPVSPMWSLWS 84

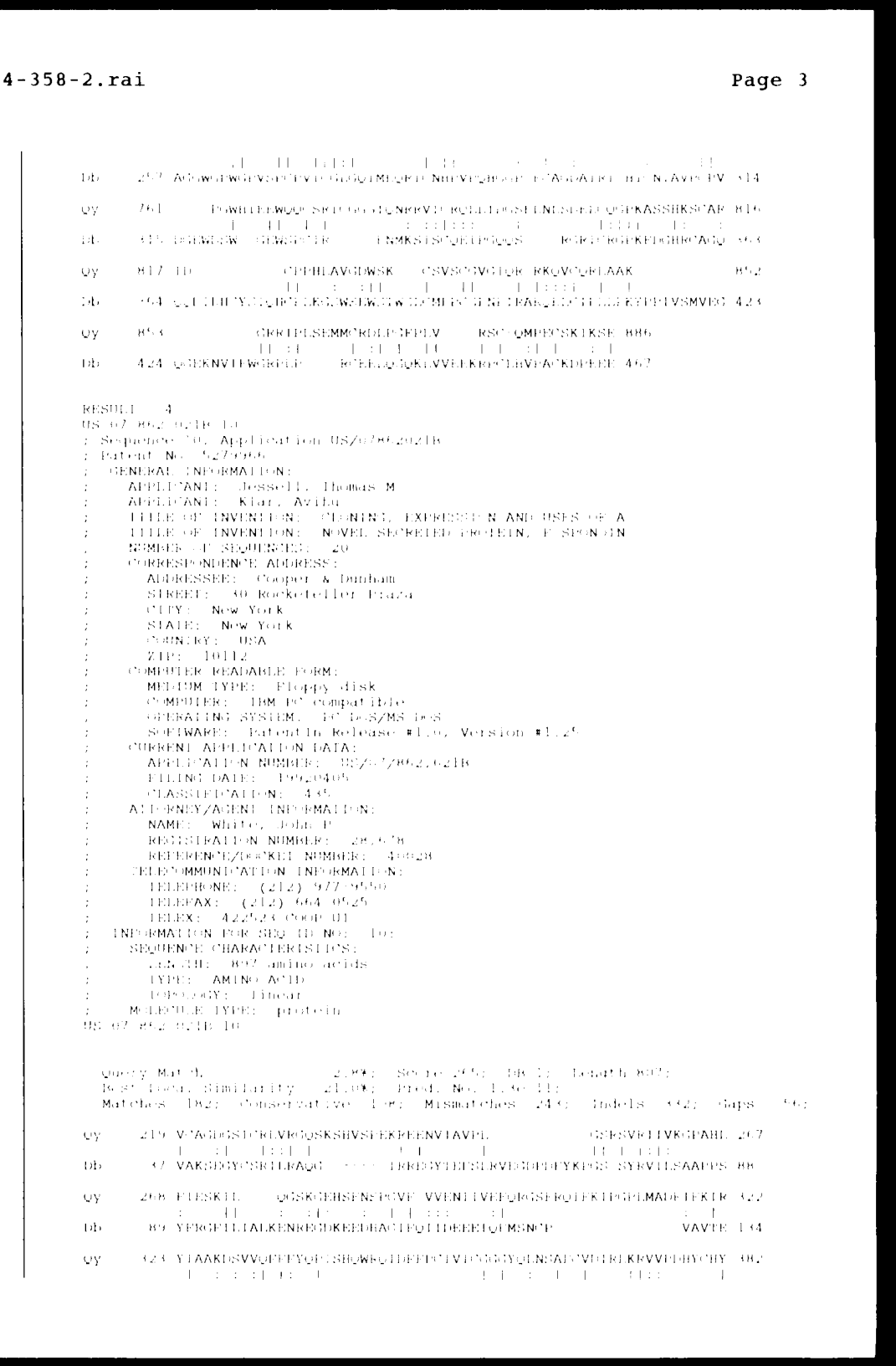
Qy 480 EWSQCIVLCGPGLRYFVVLCLNRLCELHVCQCNDQLLKCHLEEEVLILECYKPEEKSEVEA 119
Db 85 IWAPCSVLCSESSQLEYFRPCVQWNSQCSCKVAP 118

Qy 540 KLPWLKQVAGELEETPLAIELPCFD PWSACSLLCSGGVQVRELKCTVLLH 587
Db 119 ILEWQLQ ACEDQQCCLELMGDWSCWQPWEPCSVLCSKGIRFRPRA 165

Qy 588 TQIELPLPEEECEG PKLPIEFPCLLEACQESPASGELDLPLPEDSELIYDWEYA 641
Db 166 NHPAPKCGGHCPGQAGES EACDLQ QVCLPLPCAWAIWCEWL 205

Qy 642 PCTATCLGGHQEATAVCLHLQLQQQIVNDSLCDMVHRPPA MSQACNLEPCLP 695
Db 206 PCSACSCRGGHEP KELPSRRCSAPRLSQKPPSKPCPSCLAYEQQRCLCLPDCPV 256

Qy 700 RWHVGGWG PCSATCAGVGLQIRDVRCLHRGELPAPEEDCRDEKPHALQACNLQECPP 760



Db 257 ACGWGPWSPCSVSCGVGLQMRQPLCNNDVRQDGLPCNLVERPPD ACAGVACPV 314

Qy 761 PGWHLEEWGQPSPCSVSCGIGNRRVLCRQLLDFGSELNESEEFQGPKASSHKSCAR 816
Db 315 PSSWSGSW SEWSPCSR ENMKSESCQLEPCQQS RGRCPGPKEFDGHRCAGQ 363

Qy 817 LD CPPHLAVGDWSK CSVSCGVGIQR RKQVCQRLAAK 852
Db 364 CLLCHPYCQLRPCEGSWSEWSCSWSMEPCGNGIRARSLDCFIQSPPPEVSMVEG 423

Qy 853 GRKPPLSEMMCRDLPCPPLV RSCQMPPCSKIKSE 886
Db 424 GSEKNVTLWGRPLP RCEELQGQKLVVELKRCPCLHVEAVKDPEEE 467

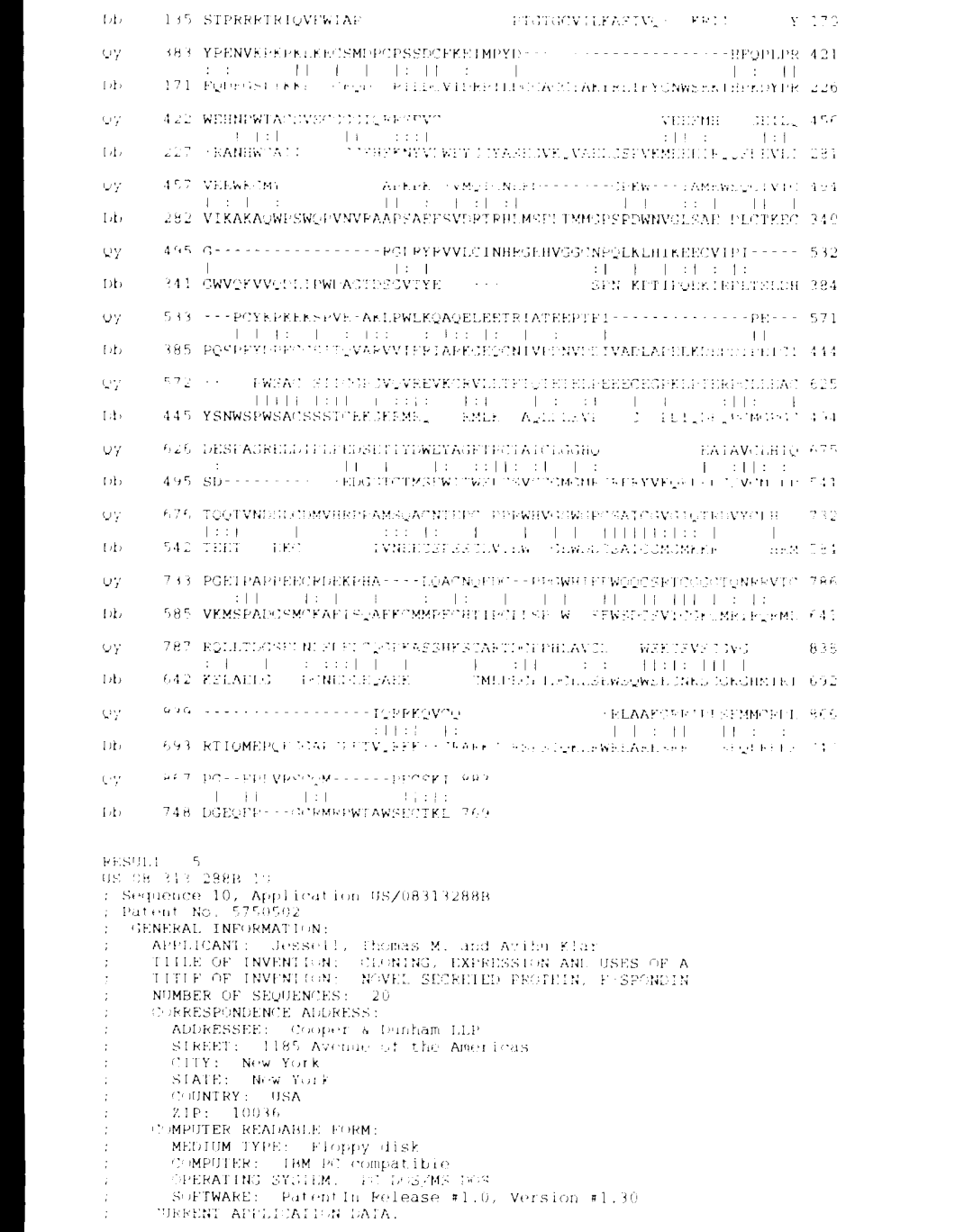
RESULT 4
US/07/862,021B/1%
; Sequence 1%, Application US/07862021B
; Patent No. 5279966
; GENERAL INFORMATION:
; APPLICANT: Jessell, Thomas M
; APPLICANT: Klar, Avihu
; TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A
; TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC DOS/MS DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/862,021B
; FILING DATE: 19920405
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 40928
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 977-9550
; TELEFAX: (212) 664-0525
; TELEX: 422523 COOP UI
; INFORMATION FOR SEQ ID NO: 1%:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 897 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US/07/862,021B/1%

Query Match 2.8%; Score 29; DB 1; Length 897;
Best Local Similarity 21.0%; Pred. No. 1.3e+11;
Matches 102; Conservative 1%; Mismatches 243; Indels 332; Gaps 56;

Qy 219 V-AGDGSICKLVRGQSKSHVSEEKPEENVIAVPL GSRSVRELIVKGPAHL 267
Db 37 VAKSDGYCYSRRLRAQC DRRSGYTEDSSDRVEGDFDEYKPGS SYRVELSAAPPS 88

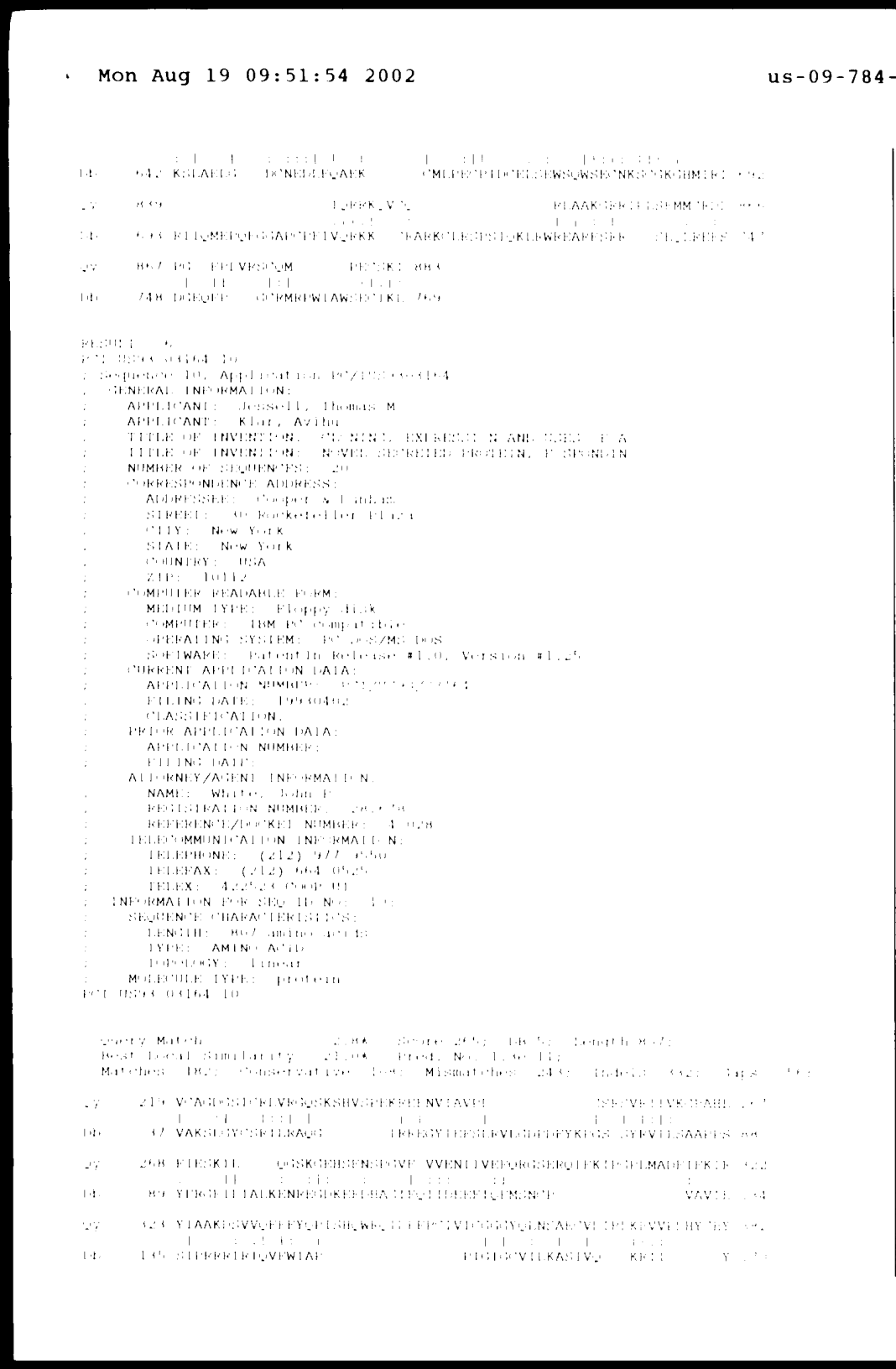
Qy 268 FLESKIL QGSKGEHSENSTSCVF VVENLIVEPQMGSEAQLEKTPGELMALGELKIR 322
Db 89 YFRGFILIALKENKRSGDKEEDRACIFQLTDEEETQFMSNCP VAVTE 134

Qy 323 YIAAKDSVVQFFEYGFCSHQWRQTDEFTCIVTSGGYGLNSAIGVDLRLKRVVFLHYCHY 382



RESULT 5
US-08-313-288B-10
; Sequence 10, Application US/08313288B
; Patent No. 5750502
; GENERAL INFORMATION:
; APPLICANT: Jossoil, Thomas M. and Avtbu Kla
; TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A
; TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/313,288B
; FILING DATE: January 5, 1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 15126-A-PCT-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION:
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 807 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-313-288B-10

Query Match 21.8%; Score 265; DB 1; Length 807;
Best Local Similarity 21.0%; Pred. No. 1.3e-11;
Matches 182; Conservative 109; Mismatches 247; Indels 332; Gaps 56;



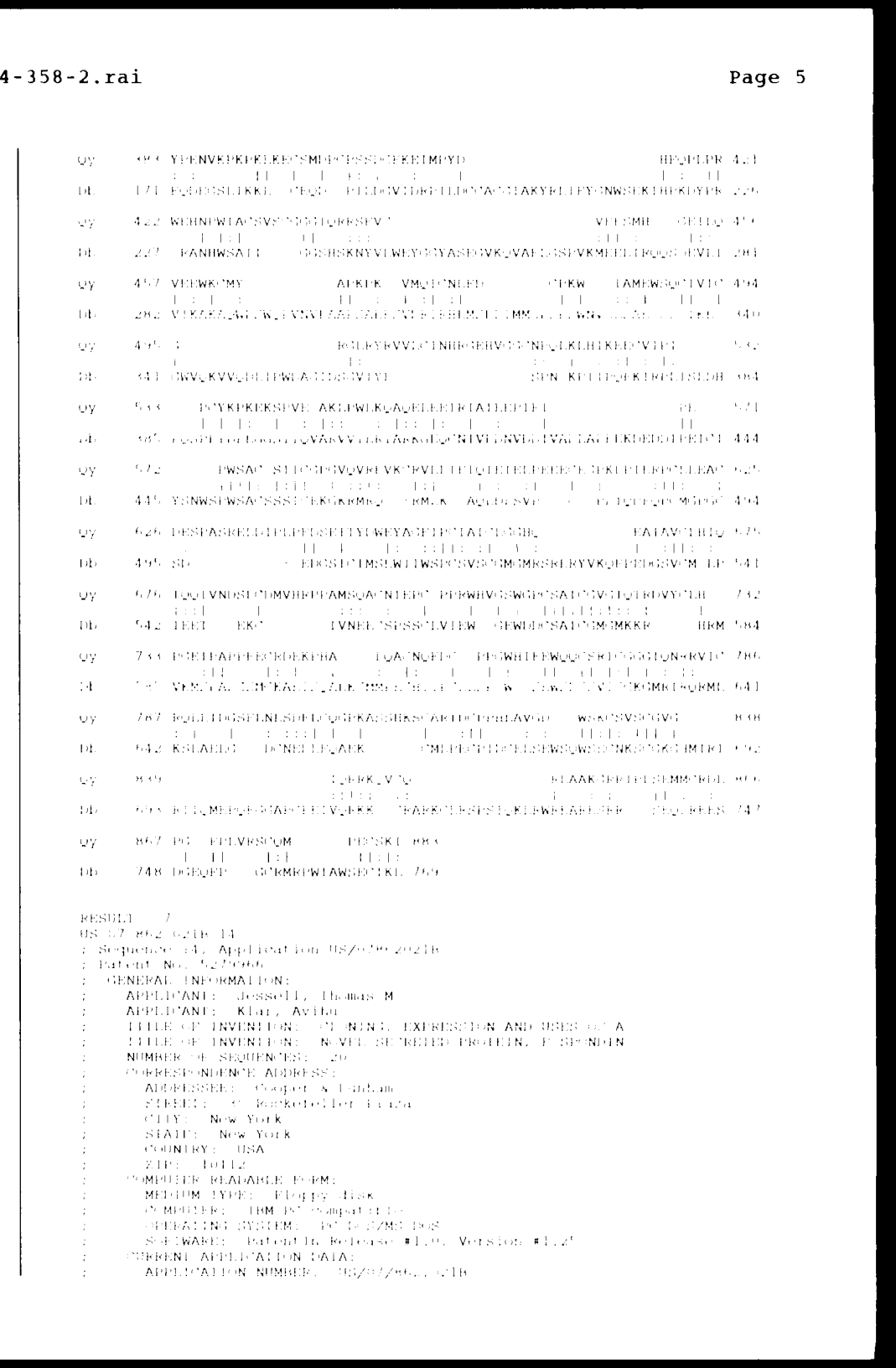
RESULT 6
PCT/US94/03164
; Sequence 10, Application PC/US94/03164
; GENERAL INFORMATION:
; APPLICANT: Jessell, Thomas M
; APPLICANT: Klar, Avihu
; TITLE OF INVENTION: MOLECULES EXPRESSED IN AND GUIDE CELL
; TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F SPONDIN
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/03164
; FILING DATE: 19930402
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 1028
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 977-8550
; TELEFAX: (212) 664-0525
; TELEX: 422523 COOP UI
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 867 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT/US94/03164 10

Query Match 2.8% Score 245; DB 5; Length 867;
Best Local Similarity 21.0% Pred. No. 1.5e-11;
Matches 182; Conservative 109; Mismatches 243; Indels 332; Gaps 7;

Qy 210 VCAGDGSICPEIVRSQMKSHVQPEPFPENVEAVPI 'SECVEIVECDABL 247
Db 67 VAKSEGYCCSRILKAQG IRESGYIPGSCERVIGDTTPYKEGS SYRVIEMAAEES 96

Qy 268 PIESKIL QWSKGGHEGENSPGVP VVENIIVEPQRGSERQIIFKIPSPLMADPIIPKIR 327
Db 97 YTRGHIIIALKENRESDKPEEMAIIPVLIDRSPTLPMGNPP VAVCL 134

Qy 328 YIAAKGVVQDFEEYLPLSHLWRLTLEPCIVDGGYYGLNEAEVTEIPLKPVVEIHY 383
Db 135 SIPRRRIRTQVPWIAP PIGIGVILKASIVQ KRII Y 175



RESULT 7
US/07/862,021B/1A
; Sequence 14, Application US/07/862,021B
; Patent No. 5270466
; GENERAL INFORMATION:
; APPLICANT: Jessell, Thomas M
; APPLICANT: Klar, Avihu
; TITLE OF INVENTION: MOLECULES EXPRESSED IN AND GUIDE CELL A
; TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F SPONDIN
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/862,021B

```
;       FILING DATE:  19920405
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  White, John P
;       REGISTRATION NUMBER:  28,678
;       REFERENCE/DOCKET NUMBER:  40028
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (212) 977-9550
;       TELEFAX:  (212) 664-0525
;       TELEX:  422523 COOP UI
;   INFORMATION FOR SEQ ID NO:  14:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  568 amino acids
;       TYPE:  AMINO ACID
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-07-862-021B-14


  Query Match              2.8%;  Score 264;  DB 1;  Length 568;
  Best Local Similarity   22.6%;  Pred. No. 9.4e-12;
  Matches  122;  Conservative   70;  Mismatches  192;  Indels  166;  Gaps   33;

Qy     408 KEIMPYDHFQPLPPWEH--NPWTACSVSCGGGIQRRSFVCVEE-SMHGRTLQVEEWKCMY 464
           |  :| |  :||   :|  :|    |:: || |   : | :|  :  ||       :|
Db     128 KPTIPQDKIRPLTSLDHPQSP----SMTPGGPIIPIARVVIERIARKGE-------QCNI 176

Qy     465 AP----------------KPKVMQTCNLFDCPKWIAMEWSQC-TVTCGRGLRYRVVLCIN 507
           |                :   :||       |    || | : || :| | |
Db     177 IPDNVDDIVADLVTEEKDEDDTPETCIY---SNW--SPWSACSSATCDKGKRMR------ 225

Qy     508 HRGEHVGGCNPQLKLHIKEECVIPIPCYKPKEKSPVEAKLPWIKQAQELEETKIATEEPT 567
                     :  :| :  : :||   ::  |          : :   | : :|  |
Db     226 ---------QRMLKAQLLLSVPCPDTQDFEDCMGP----CCSDDFASTCMMSEWIT 268

Qy     568 PIPEPWSACSTTCGPGVQVRE--VKCRVLLIFTQTETELPEEE--CEGPKLPTERPCLLE 623
             || || :|| |:||| ||          : ||:   |: |   ||:  : |
Db     269 -----WSPCSASCGMGIEVRERYVK----------QFPEDGSLCKVPTEETEKCIVNE 311

Qy     624 ACDESPASRELDIPLPEDSETIYDWEYAGFTPCTATCLGGHQEAIAVCLRIQTQQTVND- 682
           |:          |   |:| :  |:|||  | ::      |   : | |
Db     312 ECE----------PSSCIVT---EWAEWEECSATCEMGMKKR-----HRMIKMTPADG 351

Qy     683 SLC--DMVHRPPAMSQACNTEPCP-PRWHVGSWGPCSATCGVGIQTRDVYCLHPGETPAP 739
           |:|  |       |   |:| ||    |    | || ||| | :||      | |
Db     352 SMCKADTTEVEKCMMPECHIPCVLSPW--SEWSDCSVTCGKGTKTRQRMLKSPSEL--- 406

Qy     740 PEECRDE-KPHAIQACNQFDCPPGWHIEE---WQQCSRTCGGGTQNRRVTCRQLLTDGSF 795
             :| :| :   :: |  :||    : |   | :|::  | |     :  | :  :  |
Db     407 -GDCNEELELKQVEKCMLPECPISCELTEWSYWSECNKCSGKG---HMIPTRMITMEPQF 462

Qy     796 LNLSDELCQGPKASSHKSCARTDCPP-------HL---------------AVGD------ 827
           :|  |:    | |    |           ||                 : |:
Db     463 ---GGAVC--ELIVQKKKCKLEKCQKSSGNEFFHLKDAEKEEPSEKIKEDSPGEQYPVCK 517

Qy     828 ------WSKCSVSCGVGIQKRKQVCQQ-AAKGKKCPLSEM-MCRDLPGPPLVRSCQMPEC 880
                |:::|  || |||       :|     :|   |:  |:|      :|:| :  |
Db     518 MEPWIAWTKCSKECGGGIQ_      _EKRMIVKEKEKSSQLTQCPELKEL  TEACNVHPC 568


RESULT   8
PCT-US93-03164-14
; Sequence 14, Application #2/US9303164
;  GENERAL INFORMATION:
;     APPLICANT:  Jessell, Thomas M
;     APPLICANT:  Klar, Avihu
;     TITLE OF INVENTION:  CLONING, EXPRESSION AND USES OF A
;     TITLE OF INVENTION:  NOVEL SECRETED PROTEIN, F SPONDIN
;     NUMBER OF SEQUENCES:  20
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Gregory J Gore
;       STREET:  30 Rockefeller Plaza
;       CITY:  New York
;       STATE:  New York
;       COUNTRY:  USA
;       ZIP:  10112
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US93/03164
;       FILING DATE:  19930402
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  White, John P
;       REGISTRATION NUMBER:  28,678
;       REFERENCE/DOCKET NUMBER:  40028
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (212) 977-9550
;       TELEFAX:  (212) 664-0525
;       TELEX:  422523 COOP UI
;   INFORMATION FOR SEQ ID NO:  14:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  568 amino acids
;       TYPE:  AMINO ACID
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
PCT-US93-03164-14


  Query Match              2.8%;  Score 264;  DB 5;  Length 568;
  Best Local Similarity   22.6%;  Pred. No. 9.4e-12;
  Matches  122;  Conservative   70;  Mismatches  192;  Indels  166;  Gaps   33;

Qy     408 KEIMPYDHFQPLPPWEH--NPWTACSVSCGGGIQRRSFVCVEE-SMHGRTLQVEEWKCMY 464
           |  :| |  :||   :|  :|    |:: || |   : | :|  :  ||       :|
Db     128 KPTIPQDKIRPLTSLDHPQSP----SMTPGGPIIPIARVVIERIARKGE-------QCNI 176

Qy     465 AP----------------KPKVMQTCNLFDCPKWIAMEWSQC-TVTCGRGLRYRVVLCIN 507
           |                :   :||       |    || | : || :| | |
Db     177 IPDNVDDIVADLVTEEKDEDDTPETCIY---SNW--SPWSACSSATCDKGKRMR------ 225

Qy     508 HRGEHVGGCNPQLKLHIKEECVIPIPCYKPKEKSPVEAKLPWIKQAQELEETKIATEEPT 567
                     :  :| :  : :||   ::  |          : :   | : :|  |
Db     226 ---------QRMLKAQLLLSVPCPDTQDFEDCMGP----CCSDDFASTCMMSEWIT 268

Qy     568 PIPEPWSACSTTCGPGVQVRE--VKCRVLLIFTQTETELPEEE--CEGPKLPTERPCLLE 623
             || || :|| |:||| ||          : ||:   |: |   ||:  : |
Db     269 -----WSPCSASCGMGIEVRERYVK----------QFPEDGSLCKVPTEETEKCIVNE 311

Qy     624 ACDESPASRELDIPLPEDSETIYDWEYAGFTPCTATCLGGHQEAIAVCLRIQTQQTVND- 682
           |:          |   |:| :  |:|||  | ::      |   : | |
Db     312 ECE----------PSSCIVT---EWAEWEECSATCEMGMKKR-----HRMIKMTPADG 351

Qy     683 SLC--DMVHRPPAMSQACNTEPCP-PRWHVGSWGPCSATCGVGIQTRDVYCLHPGETPAP 739
           |:|  |       |   |:| ||    |    | || ||| | :||      | |
Db     352 SMCKADTTEVEKCMMPECHIPCVLSPW--SEWSDCSVTCGKGTKTRQRMLKSPSEL--- 406

Qy     740 PEECRDE-KPHAIQACNQFDCPPGWHIEE---WQQCSRTCGGGTQNRRVTCRQLLTDGSF 795
             :| :| :   :: |  :||    : |   | :|::  | |     :  | :  :  |
Db     407 -GDCNEELELKQVEKCMLPECPISCELTEWSYWSECNKCSGKG---HMIPTRMITMEPQF 462

Qy     796 LNLSDELCQGPKASSHKSCARTDCPP-------HL---------------AVGD------ 827
           :|  |:    | |    |           ||                 : |:
Db     463 ---GGAVC--ELIVQKKKCKLEKCQKSSGNEFFHLKDAEKEEPSEKIKEDSPGEQYPVCK 517

Qy     828 ------WSKCSVSCGVGIQKRKQVCQQ-AAKGKKCPLSEM-MCRDLPGPPLVRSCQMPEC 880
                |:::|  || |||       :|     :|   |:  |:|      :|:| :  |
Db     518 MEPWIAWTKCSKECGGGIQ_      _EKRMIVKEKEKSSQLTQCPELKEL  TEACNVHPC 568
```

```
RESULT   4
US-07-862-021B-12
; Sequence 12, Application US/07862021B
; Patent No. 5225069
;  GENERAL INFORMATION:
;    APPLICANT:  Bessell, Thomas M
;    APPLICANT:  Klar, Avihu
;    TITLE OF INVENTION:  CLONING, EXPRESSION AND USES OF A
;    TITLE OF INVENTION:  NOVEL SECRETED PROTEIN, F-SPONDIN
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham
;      STREET:  30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10112
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/862,021B
;      FILING DATE:  19920605
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P.
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  415-28
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 977-9550
;      TELEFAX:  (212) 664-0525
;      TELEX:  422523 COOP UI
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  802 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-07-862-021B-12


  Query Match              21.7%;  Score 250;  DB 1;  Length 802;
  Best Local Similarity   20.0%;  Pred. No. 9.5e-11;
  Matches  179;  Conservative  103;  Mismatches  278;  Indels  208;  Gaps   53;

Qy     205 DRQLQSSNAKEENCSVEVAGRSSLPSLVRSQSKSHVSPEKREENVIAVP  253/VPEIV 261
           | |   ||  :  | |:      |:|| |       :   |  :      ' |:::  :
Db      26 DEILEKAAKSE  GYCS      RILRAQGIRREGYNEPSLRVEGDEEFYKPCNSYEVIL 77

Qy     262 KGPAHELPTEGKEI     QVSKCPHGSPNSPCVP VVENILVEPQLPCSPPQIPKEIPGCMAL 320
           :  ||    :| :|:   :  | | :::    :|                          :
Db      78 SAAIPAYEPCPTLLTALKPCKECDKEEDHAGIPVLYDEEEQGPMSNCT              124

Qy     321 PIPKIPYIAAKQVVQPEIY QCLGLQWRQIIPEFVD ZGGYQLNSAEVCLKS        371
           : | |  : :: :| | :  |                 | |  :  |  |  :
Db     125  VAVPSPPPPPPIPLVEWSS                PLDDVCLKAVVQRDIYLQ  172

Qy     372       KRVVPDHYCHYYFENVKPKELKKGCSMDGKVSGSVSPEKEIMPY       418/PPIP 430
                 ||:           | :      ::: |       : |           :
Db     173 EGSLLKRDFPQLGA    SEGVCLKP      IIPCCAVGIAKYRLIFVSNWSEKIHPKLDPT 224

Qy     421 RWEHNFWIACSVSP             SS    GIGR     KSPVCVIELSMG    SCSLVEL 479
           |   | |:|   |                |    |:::      | | :|| :      |:    :
Db     225 RPINHWSAIDGSHSKNYTLWYVGGYASEGVKQVAELGSPVKMPSETPQQQTLEVLIVIK 284

Qy     480 WKCMYAIPKIKVMGTCNLPSC                KWIAM      EWS        QCIVLC 534
           |  :     |  |:  |                     ::|      :|:        :| |  |
Db     285 AKAQW     FAWQPLNVPAAIGALCVSFFFLISIIIGICIGFCKWNVKGNLAANDQKSVNTLNN 344

Qy     495 G                   RCLRYRVVLCINHRGEHVGGCNPQLKLHIKEPCVTPC        542
           |                     : :|                 :  |  :  :  : :|
Db     345 GWVQKVVQKLFEWDAGTDGSGVIYE                  SEN-KPIVPQEKIRPLISLDH 399

Qy     543     PCYKKPEEKSPVEAKDLWEK_AQECLEEIKCAI  PEPIEIH                   591
               | | |:  |          :|    :  |||   |:  |:|:
Db     400 PQSPEYSPPSGS          EKLVAPVVLLPGIAKKCGQCNEVIENIICLVAIGLAPCEKPE 459

Qy     592                PWSAC STLSSPGVQVREVKCPKVILLIFTQIEIKCPEEPCPGPKLPIF 641
                          ||||| |:||  | ::::   |:|    :  :  ::    :  |
Db     460 IGEEEECTTSNWSFWSACSSSCEKGKRMRQ   KMLK  AQLDLSVF   C  PDTQQP 511

Qy     642 RPCIERAVDLSFASNELLLPLIIGSFTYDWGYAGEGPCIAIC PPGHQLATAVCIHIQS 701
           :::|   | :                |:  |   ::  :::|: :: :,    |
Db     512 QPCMGPGPSD              PDGSICMMSDWIIWSFCSVSCGMGIRSREPYV      K 564

Qy     702 PQIIVNDSLCPMVHKEEAMSQAC NIEPCPE        RWHVGSWSPCSATCGVGIQQRDVY 759
           |   : :|| :   |   :: |  | | |         |  | |  | |||  : :|
Db     565 QPPEPGSMCKV   PPEELEEKCIVNEPCPSDCELVEEW   NEWEP CSASCGIGMKQR     576

Qy     760 CIHP  GEPPAPPEPCRDEKPHA    IQACNQFDC  PPCWHIEEWQQCSKPCGGGIQNR 781
           :     |||    |: | :    :   |: |  :|  |  :| |  :  | |:| : :
Db     577  HSRCEMSEEALSLMEMAALIGAGA RMIA HLCI CSGF W  CSWSCSVICGKGMRI 631

Qy     782 REVICRQLLIDGSPLNESDEDCQCPKASSHKSCARIPCPPDIAV   GEWSKCSVSCGVG 838
           ::    | | :       : :|||  |  |      |  :||    :    |:|:| ||| |
Db     632 RL   RMLKSAAIDGSPNRSDLGALK       MLPLCIIICSELEWSQWSECNISCGKG 682

Qy     839  IQFRECVCQLEAAKCRKCP  ICDMPSCRSLPCPPLAVRCSQMLEGSEKDGMKIKDGEGC 895
            : | :       |   |        :|      |||                    :|
Db     683 RMIRLRMKIKDGSLASLELVQGKLCP        VEKC                   RG 716

Qy     896 PQFLSVQRVYIQIREEKR 913
           |  :   : : : ||:|:|
Db     717 PGM  EKRPWELAPERRR 734


RESULT  10
US-08-313-288B-12
; Sequence 12, Application US/08313288B
; Patent No. 5750502
;  GENERAL INFORMATION:
;    APPLICANT:  Jessell, Thomas M. and Avihu Klar
;    TITLE OF INVENTION:  CLONING, EXPRESSION AND USES OF A
;    TITLE OF INVENTION:  NOVEL SECRETED PROTEIN, F-SPONDIN
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham LLP
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/313,288B
;      FILING DATE:  January 5, 1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P.
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  415-28 A PCT-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 278-0400
;      TELEFAX:  (212) 391-0525
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  802 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
```

```
;    MOLECULE TYPE:  protein
US-08-413-288B-12


  Query Match              2.7%;  Score 256;  DB 1;  Length 802;
  Best Local Similarity   20.9%;  Pred. No. 6.5e-11;
  Matches  179;  Conservative  103;  Mismatches  278;  Indels  298;  Gaps   53;

Qy     205 DPQTCSNAFIPNCVVCAGCYSTCPLVEVQSESHVSPEPEPEENVTAVP---IPSPSVPETV 261
           |  |   || :  | ||       |::| |         :    |   |         | |:|:
Db      26 DETILEKAAKSE  GYCS      EILEAQCTRRGCYNEFSLEVEGRPEFYKPGNGYRVTL 77

Qy     262 KGPAHLFIESKTL    QGCPGGEHSPNFPCVF VVENTTVEPQRGSERQTPEKIPGPLMAD 316
                 :    ||     :| :|:   :  | | :::    :|
Db      78 SAATPAYFRGFTLIALKEGKEGDKEEDHAGIFQTIDEEETQFMSNCP--  -----     -- 124

Qy     317 FIFKTPYTAAKDSVVQFFFVQPISHQWRQTIDFFPCTVPGGGYQLNSAECVPIRL----- 371
           :  |  |  : : :| |: |          | | : |  | |  |:
Db     125 -VAVTESTPRRRTIRIQVFWTAP---------------PTGTGCVILKASIVQKRTIYFQD 168

Qy     372       KRVVPEGYCHYYPLNVKFKPKLKLGGSMIFPCPSSEGFKEIIMPY       DPPQPLP 420
                 ||:          | |  ||      ::| |          :  |       | :  |
Db     169 EGSLTKRICEQDSA---SEGVTDKP-----ILPCACGTAKYRLIFYGNWSEKIHPKDFP 220

Qy     421 RWEHNPWTAGSVSC----------GG----GIQR----RSFVCVEESMH  --GEILQVEE 459
           |   | |:|   |             ||     |:::      | | :|| :     :|: | :
Db     221 R RTNHWSATIGSCHCRANYTLWLYGMVAAEDVRQVALLGSTVRMTIIPQQGQLVEIVEK 279

Qy     460 WKCMYAPKIKVMQICNLFDCP---------- --KWIAM----EWS--   ---QCTVTC 494
           |  :        |  |:  |              :: |     :|:        || |
Db     280 AKAQW----PAWQPLNVRAAPSAEFSVIRHRHLMSFLIMLIGPSPDWNVGLSAEDLCTKDC 335

Qy     495 G   ---   --- ----RGLRYRVVECINHRGEHVGGCNPQLKLHIKEEGCVIPI  --- 532
           |                 |: |              :|  |   : :| : |:
Db     336 GWVQKVVQPFDIRWTACIGSCVTYF ----------SRN-KPIVPQEEIRPLTSLDH 379

Qy     533 ---PCYKPKEKSPVEAKLPWLFQAQELEETRIAT--EEPIPIPE---------------- 571
               | | |:  |        :|    :   |||   ::  |:|:
Db     380 PQSPEYIDPEGGS--------IFLVAKVVLGSIAKKGLQCNEVGNIIGSVALLALHLKEL 431

Qy     572  ---     -----PWSAC--SITCGPGVQVRFVKCRVLLTFTQTETELPEEECEGPKLPTE 617
                      |||| |:||   | |:|:    |::    | :  :|    |  |
Db     432 LITPEETTIYSNWGPWSACSSSTCFRSPMKQ   RMIP  AQIIISVE   G  PTTQDT 491

Qy     618 RPCLLEACDESPASPPLDIPLPEDSETTYDWEYAGFIPCTAIC-LGGHQEAIAVCLHIQT 676
           ::|:   | :              || |     ::  ::|:: :| :|       |
Db     492 QPCMGPGCSD------------EDGSTCMMSDWTIWSPCSVSCGMGIRSRERYV-----K 534

Qy     677 QQTVNDSLCDMVHRPPAMSQAC-NTEPCPD------KWHVGSWGPCSATCGVGIQTRDVY 729
           |   : |:|  :   |   :: |  | | |       |  | |  |||:|| |:: |
Db     535 QFPEDGSMCKV---PTPFTPKCIVNEECSPSSCLVTEW--GEWDECSASCGTGMKRR--- 586

Qy     730 CLRP--GRTPAPPEECRDEKPHA----LQACNQPDC--PPGWHIEEWQQCSRTCGGGTQN 781
           |     |||    |: |  |    :  |:   |  | |   ||  || ||| | :
Db     587 --HRMIKMIPAPGSMCKAKIIEAEKCMMPECRIIPCLLSP-W--SEWSPCSVTCGKGMRI 641

Qy     782 RRVTCR-LLDGSTVNT-SFCLQERAESHFIAESCVMSFIHGAV   SWAECSVSCGV 838
           |:   | | :      : :::| |  |       |   :||     :     ||:|: ||| |
Db     642 RQ---RMIKSAARDGDCNSELEQAEK------CMLPECRIIPCLLEWQQWSECNTSCGKG 692

Qy     839 -IQPRKQVCQPIAAKGRRIP--ISEMMGPDIDCPEPIVPFSCQMDPCSRIPSEMFTKIGEQG 895
            : | :        |  |  :    ||       || |              :|
Db     693 RMIRIKMIKIEPQWGGIACPELVQRIKCR-------VRKCL-   --    --------RG 718

Qy     896 PQILSVQRVYIQIREEKR 914
           | :   :| : : :|:::|
Db     719 PGM--EKRRWKEAKEKRK 744


RESULT  11
PCT-US93-03164-12
; Sequence 12, Application PCT/US93/03164
;  GENERAL INFORMATION:
;    APPLICANT:  Jessell, Thomas M
;    APPLICANT:  Klar, Avihu
;    TITLE OF INVENTION:  CLONING, EXPRESSION AND USES OF A
;    TITLE OF INVENTION:  NOVEL SECRETED PROTEIN, F-SPONDIN
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham
;      STREET:  30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10112
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Patentin Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/03164
;      FILING DATE:  19930402
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  40028
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 977-9550
;      TELEFAX:  (212) 664 0525
;      TELEX:  422523 COOP UI
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  802 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
PCT-US93-03164-12


  Query Match              2.7%;  Score 256;  DB 5;  Length 802;
  Best Local Similarity   20.9%;  Pred. No. 6.5e-11;
  Matches  179;  Conservative  103;  Mismatches  278;  Indels  298;  Gaps   53;

Qy     205 DPQTCSNAFIPNCVVCAGCYSTCPLVEVQSESHVSPEPEPEENVTAVP---IPSPSVPETV 261
           |  |   || :  | ||       |::| |         :    |   |         | |:|:
Db      26 DETILEKAAKSE  GYCS      EILEAQCTRRGCYNEFSLEVEGRPEFYKPGNGYRVTL 77

Qy     262 KGPAHLFIESKTL    QGCPGGEHSPNFPCVF VVENTTVEPQRGSERQTPEKIPGPLMAD 316
                 :    ||     :| :|:   :  | | :::    :|
Db      78 SAATPAYFRGFTLIALKEGKEGDKEEDHAGIFQTIDEEETQFMSNCP--  -----     -- 124

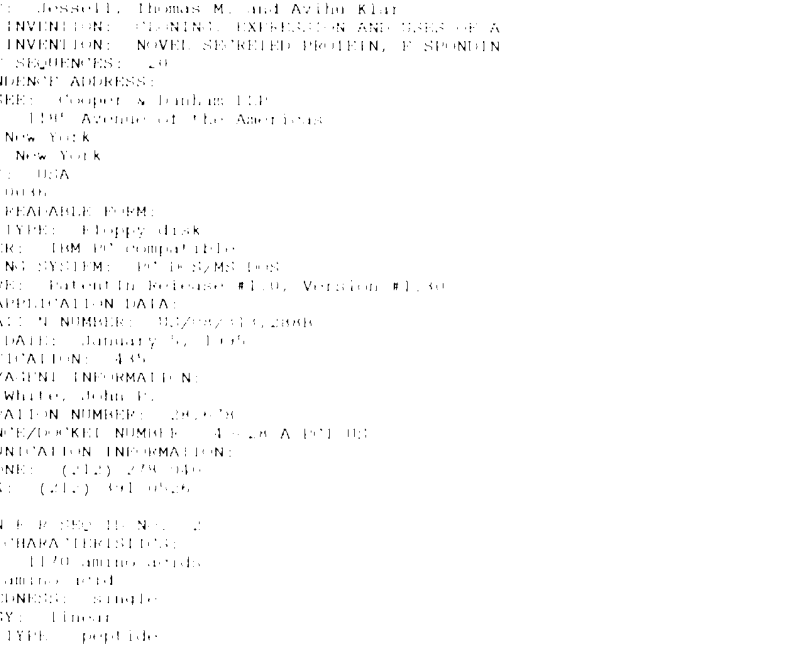
Qy     317 FIFKTPYTAAKDSVVQFFFVQPISHQWRQTIDFFPCTVPGGGYQLNSAECVPIRL----- 371
           :  |  |  : : :| |: |          | | : |  | |  |:
Db     125 -VAVTESTPRRRTIRIQVFWTAP---------------PTGTGCVILKASIVQKRTIYFQD 168

Qy     372       KRVVPEGYCHYYPLNVKFKPKLKLGGSMIFPCPSSEGFKEIIMPY       DPPQPLP 420
                 ||:          | |  ||      ::| |          :  |       | :  |
Db     169 EGSLTKRICEQDSA---SEGVTDKP-----ILPCACGTAKYRLIFYGNWSEKIHPKDFP 220

Qy     421 RWEHNPWTAGSVSC----------GG----GIQR----RSFVCVEESMH  --GEILQVEE 459
           |   | |:|   |             ||     |:::      | | :|| :     :|: | :
Db     221 R RTNHWSATIGSCHCRANYTLWLYGMVAAEDVRQVALLGSTVRMTIIPQQGQLVEIVEK 279

Qy     460 WKCMYAPKIKVMQICNLFDCP---------- --KWIAM----EWS--   ---QCTVTC 494
           |  :        |  |:  |              :: |     :|:        || |
Db     280 AKAQW----PAWQPLNVRAAPSAEFSVIRHRHLMSFLIMLIGPSPDWNVGLSAEDLCTKDC 335

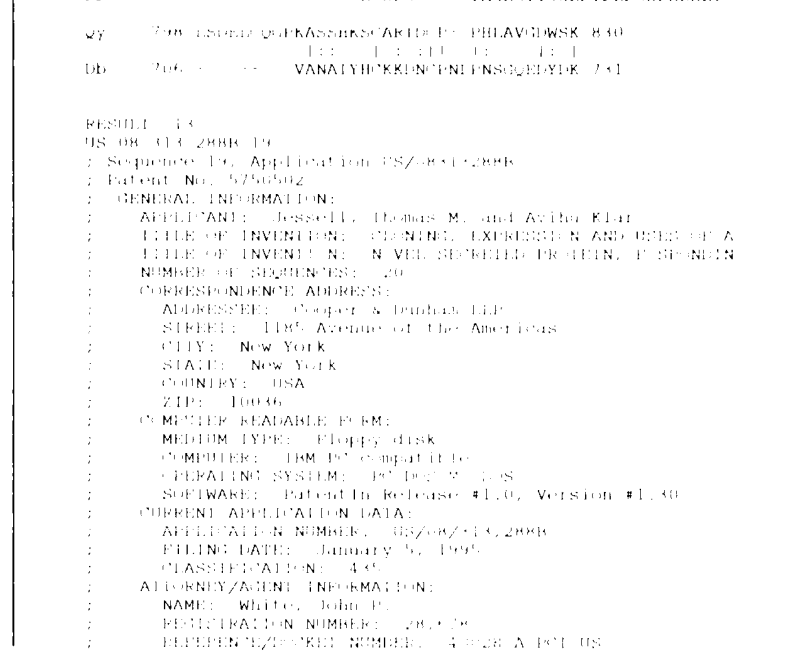
Qy     495 G   ---   --- ----RGLRYRVVECINHRGEHVGGCNPQLKLHIKEEGCVIPI  --- 532
           |                 |: |              :|  |   : :| : |:
Db     336 GWVQKVVQPFDIRWTACIGSCVTYE ----------SRN-KPIVPQEEIRPLTSLDH 379
```



RESULT 12
US-08-814-208B-20
; Sequence 20, Application US/08814208B
; Patent No. 5750502
; GENERAL INFORMATION:
; APPLICANT: Jessell, Thomas M. and Arthur Klar
; TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A
; TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/814,208B
; FILING DATE: January 5, 1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 415.28 A PCT US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; TELEX:
; INFORMATION FOR SEQ ID NO: 20
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1170 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide

US-08-814-208B-20

Query Match 2.6%; Score 240; DB 1; Length 1170;
Best Local Similarity 21.4%; Pred. No. 2.2e-11;
Matches 110; Conservative 94; Mismatches 165; Indels 185; Gaps 20;



RESULT 13
US-08-814-208B-19
; Sequence 19, Application US/08814208B
; Patent No. 5750502
; GENERAL INFORMATION:
; APPLICANT: Jessell, Thomas M. and Arthur Klar
; TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A
; TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/814,208B
; FILING DATE: January 5, 1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 415.28 A PCT US

```
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (212) 278 8400
;       TELEFAX:  (212) 391-0526
;       TELEX:
;   INFORMATION FOR SEQ ID NO:  13:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1172 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  peptide
US-08-414,288B-13
```

```
  Query Match              2.4%;  Score 249;  DB 1;  Length 1172;
  Best Local Similarity   23.8%;  Pred. No. 1.9e-69;
  Matches  114;  Conservative   56;  Mismatches  172;  Indels  152;  Gaps   35;

Qy     395 PCSMIPPCPSS----GCPKFIMPYIHPQPLPPWEHNPWIACSVSCGGGIQRRSFVCVEESM 450
           ||   |||     || :   |          |    || ||:||  | |:|    |
Db     369 EC----CPSCLHSVIGEEGWSP-------  WAE- WTQCSVTCGTGIQFRSRC-      408

Qy     451 HGEILQVEPWKCMYAPKPKVM-QTCNLEDCPKWIAME--------WSQCTVTCGPGLRYR 501
                 |    |:    | :  : |:|  |   |  :          || |:|||| |   |
Db     409 -----DVISNICL   GESIQYRACSLSKCPIKEKQKSGWLHWSFWSSCSVTCGVGNITR 460

Qy     502 VVLQCNHRGDEVGGCNEQLDPLHIPLEPCVIDPCPEEIPEPEPVEAE LWWLPLALLLLLLL 561
           : || |    :|| | :      : |    ||        |:: :  ||
Db     461 IRLC-NSPVPQMGGPNCKGSGRETEACQ-CAPC-------PIDCRWSPW----------- 500

Qy     561 IATEEPIEIPEPWSACSTICGPGVQVRSEVKCKVLIPTQTEIELPEEECEGP  --KLPT 616
                 ||||| ||  |:: |   |          | :   : | |     ::
Db     501 ----------SPWSACIVTCAGGIRERTRVC------NSPEPQYGGKACVGDVQERQMCN 544

Qy     617 ERPCLLEACDESPASKFLDIP      -LPEDSETTYIDWEYAGFIPCTATCIGGHQEAIAV 670
           :| | :: |  :|       |        |: |     |  || |       | | | :
Db     545 KRSCPVDGCLSNPC-----PPGAQCSSFPDGS-----WS-CGFCPVCFLGNGTHCEDLDE 594

Qy     671 CIHIQTQQIVNDSLCLMVHRPPAMSQACNTP------PCPPRWHVGSWGPCSATCGVGIQ 724
           | :          :|   : |   : ||        |||||:          :   ||||::
Db     595 CALV  ---  PDICFSISKVP--  PCVNTQPGFHCLPCPPRYR--  ---GNQPVGVGLE 647

Qy     725 ---IRDVYCLHPGETPAPPEEPCRDEKPHALQACNQP-DCPPGWHIEE-WQQCSRTCGGGT 779
               |  |       :|  |:| | |   |::  :|    :  :|    | |
Db     648 AAETEKQVC   -   EFENGKD KTH-  NCHKRAECTTYGHELSDEMTKVE--QQTGY 684

Qy     780 QNPRRVICRQLITDKSPLNESRECQDPPASSHFSCARTPWQ--PHLAVGHWSPCSV--SC 835
           : | :      : ||:  |      |: |  | : :||  |:    |: |  :  :|
Db     685 AGDGLICGEDSDLDGWPNLN--LVCAINATYH  CIKDNCPHLPNSGQEDFDKDGIGDAC 740

Qy     836 -----GVGIQRRKQVCQRL 849
                 |:   |  || :
Db     741 DDDDDNDGVTDEKDNCQLL 759
```

```
RESULT  14
US-08-752-307B-14
; Sequence 14, Application US/08752307B
; Patent No. 5952171
;  GENERAL INFORMATION:
;    APPLICANT:  McCarthy, Sean A.
;    APPLICANT:  Guritz, David T.
;    APPLICANT:  Levinson, Douglas A.
;    TITLE OF INVENTION:  METHODS FOR IDENTIFYING GENES
;    TITLE OF INVENTION:  ENCODING NOVEL SECRETED OR MEMBRANE ASSOCIATED PROTEIN
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson, P.C.
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  US
;      ZIP:  02110-2804
```

```
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  Windows95
;       SOFTWARE:  FastSEQ for Windows Version 2.0
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/752,307B
;       FILING DATE:  19-NOV 1996
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Meiklejohn, Ph.D., Anita L.
;       REGISTRATION NUMBER:  35,283
;       REFERENCE/DOCKET NUMBER:  09401/020001
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  617-542-5070
;       TELEFAX:  617-542-8906
;       TELEX:  200154
;   INFORMATION FOR SEQ ID NO:  14:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  640 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-08-752-307B-14
```

```
  Query Match              2.3%;  Score 213.5;  DB 2;  Length 640;
  Best Local Similarity   21.1%;  Pred. No. 7.9e-08;
  Matches  134;  Conservative  104;  Mismatches  223;  Indels  173;  Gaps   35;

Qy     871 LVEASQLMPECDEAIRALDRLELSGQQIQTLDVAFVSIQSLERLRNLFSSRAVIL ENLS 929
           |: : :::  | |:|||:          |  :: | |                 : |:: | :
Db      22 LISAIEVPLDSNIQSELPQ       PFTITKQSV -                 KDYIVDPKDN 59

Qy     930 VLIPQPVPPEQVSFSQWLFKFDPSTQNSRMCSII---KSPSLF---HGLAAFG--GVYPC 982
           : |:|  :          | ::|:    :|   ::   :|||:| |  ||   ||   | |:|
Db      60 IPIECPAFGNPVPEISWIRNGQPIRNVAGTEVSMMERGCSILVIDEHGGGSRDDYEGEYQC 119

Qy     983 IAGSAQETVVLKLIGIDNRLIARPALREPMREYPGMDHSEANSLG-------------VI 1029
           | :   | :   |   :  ::|  | |  :   |  |  :   |               :
Db     120 FANNATGTALSSRI   HELGVGKSFL VFRIKAVSVTLKVLGAPSKLQGNFTDGLLEVIF 179

Qy    1030 W        -HEMEQMGGNNEN ::PLLL        DRIRNGSFLRALLSRGSNSAGSINS 1072
           |        |: :::   :| |||  :         |:  |  |      | :::   |
Db     176 WMSSSMEPIHQVPFVSQVQNCDEYPTNVMIQARTDYSGNAPF      RPTHPTPQPENP 229

Qy    1073 WPLPNPSPTAAVPQCAYQMKIAQFDPFTRNPYEQLMKTCPFVSTEI --        ASQLIY 1122
           : ||       ||      :|:     :|| : :    |::              :|||::
Db     230 YILK-------VKTKKPRNETS     LENHTLMYSARGVTETIFSFMYFYGISSSQMVL 277

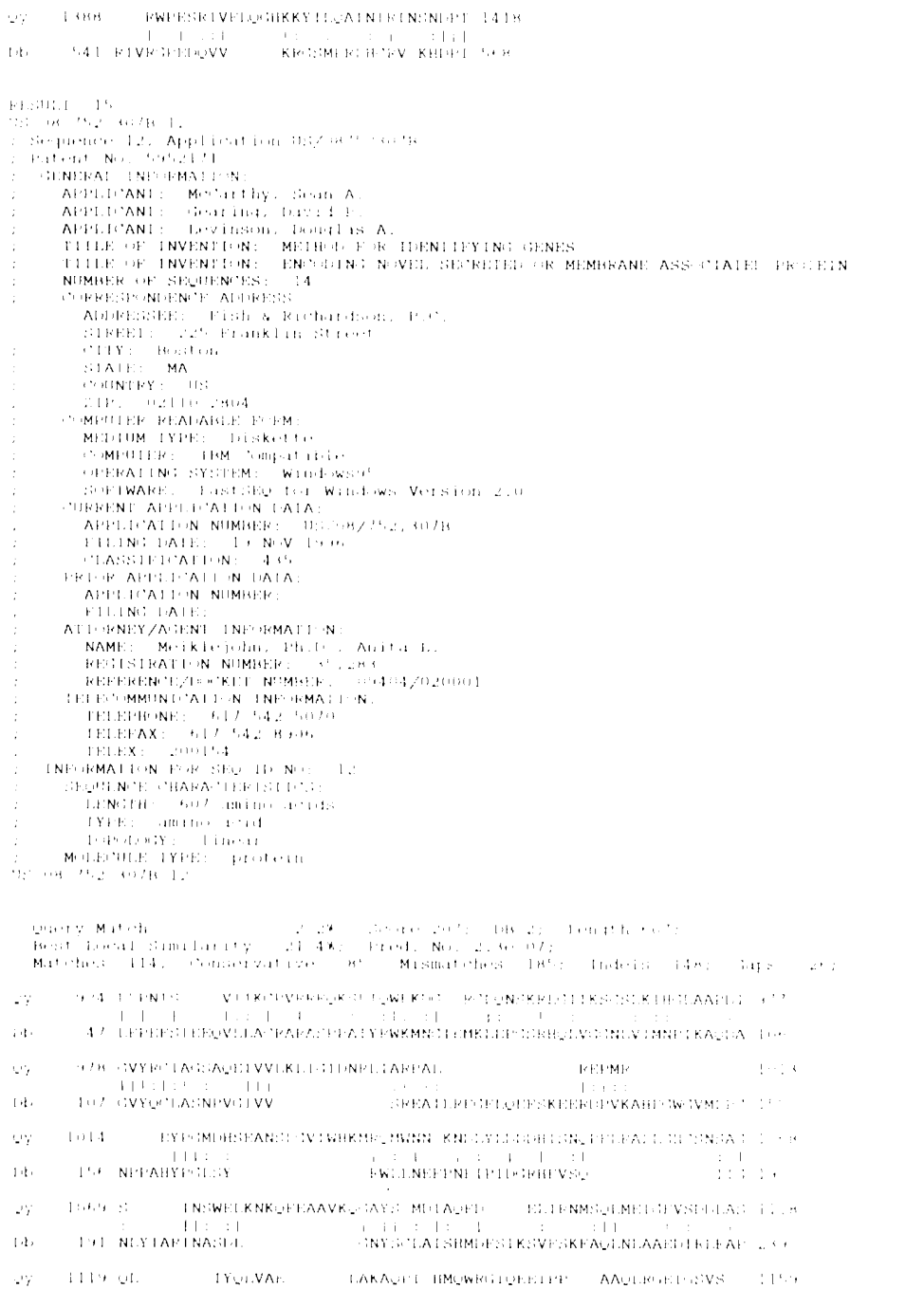
Qy    1123 Q   LVAELAARQPT HPQWFIQL---IAAQIPGLCNVQSLDLAERGCGLTPLGL 1177
           :   |:  :|   |  : :|     | |       |:|  :   : |      :  |
Db     278 RSVCLLGLCIAQVPAILRWIEKCDIS      ASR  RLLHNRAQGLSNVSEEDSGE 331

Qy    1178 VIMSQTQFFSIGFNFTINFPICNTVYIIEETHVINIICQLITPSEATYIWTKDSTLL    1234
                |    |   ||: :    |      : :     :: | |        || |:
Db     332 YFCLAGNKMG STRHTLGVRVKAAFYWLIEPSNL     ILAFSE       IAGKLVCKA 378

Qy    1235     QLPVFIIISIGRIVGLV    NLIKKEQS              ITEQSVANRLLSG 1272
                :|||:: :::|     |:    |||:|: |             :|:|: :|   |
Db     379 NGNPKFSIQWLVNGE   FIKSSPFNFSKEVAGGTIVFKDTQIGSSAVYQCNASNEHG-- 433

Qy    1273 VFSSEVFVAPAPV IISVPFNITKEEHNILSVVVGGIVEAALGANVTLPCPVKGVPQPNI 1331
                 | | | :| |   | :  : :|          : || | | | :
Db     434 -----YILANAFVSVILDVFPRTLAPRNQLIKVIQYN- ----  RTRILDCPFFGSPIPTL 480

Qy    1332 TWLKRG-GSL  SGNVSLIFNGSLLIGNVSLFNEGTYVCIATNALGKAVATSVLHLLER- 1387
           | | | ::   ||     |||| :     |::| | :|||| |||  |   |  : :
Db     481 SWFENGQQNMLIGGNYKAHENGTLDMDMARKIGLSTTGVAINILGKVIAQVELIVEDDPI 540
```



RESULT 15
US-09-784-358-2
; Sequence 12, Application US/09784358
; Patent No. 5952171
; GENERAL INFORMATION:
; APPLICANT: McCarthy, Sean A.
; APPLICANT: Goerlitz, David F.
; APPLICANT: Davidson, Douglas A.
; TITLE OF INVENTION: METHOD FOR IDENTIFYING GENES
; TITLE OF INVENTION: ENCODING NOVEL SECRETED OR MEMBRANE ASSOCIATED PROTEIN
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/784,358
; FILING DATE: 16-NOV-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Meiklejohn, Ph.D., Anita L.
; REGISTRATION NUMBER: 37,283
; REFERENCE/DOCKET NUMBER: 07684/020001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 12
; SEQUENCE CHARACTERISTICS:
; LENGTH: 607 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-784-358-12

Query Match 27.2%; Score 207; DB 2; Length 607;
Best Local Similarity 21.4%; Pred. No. 2.6e-07;
Matches 114; Conservative 85; Mismatches 185; Indels 134; Gaps 27;

Qy 934 EYPNIS VEIKPVVKKQKSILQWEKPS PVSLQNPKMKPLTLKSPSLKIHSLAAPLL 977
Db 47 LPPEPSIEEQQVLLATPAPASPPAEYPWKMNSTEMKLLPSQRHQAVSCPNLVIMNPIKAQLA 106

Qy 978 GVYPCTAGSAQEIVVLKLESIDNPLTARPAL PEPMK 1013
Db 107 GVYQCTASNPVGIVV SPEALILPPGPTLQEPSKEEKSPVKAHPSWGVMG 157

Qy 1014 EYPSMDHSPANSSVIWHKMKPLPWNN KNEGYIIGSHELSN EFELPAILSLSNSA 1069
Db 158 NPPAHYPGLSY PWELNEPPNEIPTLPGMHPVSQ 191

Qy 1070 S INSWELKNKQFEAAVKQASAYS MDLAQFD ECLENMSQLMEIGFVSEFLAS 1123
Db 192 NEYIAPLNASDL SNYSCPLAISRMDFSLKSVPSKFAQLNLAAEDLKLEAF 242

Qy 1124 QL IYQLVAE LAKAQFT HMQWRGTQEEIPP AAQLRQELQSVS 1169

Db 243 SIKKARFDAEIYALVSQLVELDPAFQNHVPMLKWPKVLQSLSPQWLTAEPLQLPSVSPE 299

Qy 1170 QSSHAKNE PKLLFKPKQLVLMPLSQLPPSTISLENKLLNSRCTNLVYLPKPLEVINLL 1214
Db 300 DEGLYECEAENSKGRDLVQGF LIVLAQLPEWLKVISLLEALQGSN LSWG 347

Qy 1215 PLLLLVSEAIYLWIPPPSLL QLPVKELLLPSDKLVLQNPLPKPLQYLYLPSVANHLGSD 1274
Db 348 PAAAGKPPPLYWWLRNGSPLAQLNPVLVL AGSLELSQELSLEDSGMYGPVAENKHGT 403

Qy 1275 VESSSVLYAIDAPVILSVEGNLTKLEIPNHLLSVVVGSTVLEAALSANVTLPFVKGVPQLPNLT 1334
Db 404 -IYASAE LAVQALAPDFKLN---PVRRLIPAARQGELLIPPQPRAAPKAVVL 451

Qy 1335 WLKKGGSL SGNVSLLIPNGSLLLQNVSLENEGIYVCIATNAIGKAVAISVL 1382
Db 452 WSKGIELLVNGSKVIVPLGSLLLENGSGEGKYFLALNPMGKANSLHL 498

Search completed: August 19, 2002, 07:45:03
Job time: 7177 secs

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

Matrix: nucleic search, match=1,mis=-3

Run on: Aug 19, 2002, 02:51:24 ; Search time 95.2,97.5 Seconds
(without alignments)
1002.166 Million cell updates/sec

Title:
Perfect score: 5070
Sequence:

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: em_estba:*
2: em_estbo:*
3: em_estfu:*
4: em_estin:*
5: em_estmu:*
6: em_estov:*
7: em_estpl:*
8: em_estro:*
9: em_htc:*
10: em_est1:*
11: em_est2:*
12: em_htc:*
13: em_htg:*
14: em_gss_fun:*
15: em_gss_inv:*
16: em_gss_pro:*
17: em_gss_vrt:*

Post No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	603.4	11.9	606	10	BG750735	BG750735 bg22b924
c 2	474	9.3	185	9	AA787148	AA787148 ; ; 0
3	409.6	8.1	846	11	AK020361	AK020361 Mus musculus
4	401.4	7.9	425	10	BE767952	BE767952 CM1-CN006
5	394.2	7.8	1000	10	BM476141	BM476141 AGENCOURT
c 6	293.6	5.8	338	9	AA006029	AA006029 cp06a07.s1
7	293.6	5.8	622	9	BB663795	BB663795 BB663795
8	281.8	5.6	800	12	AQ743958	AQ743958 BE-55b4 A
9	278	5.5	850	10	BE808802	BE808802 cc15c08.2
10	274	5.4	619	9	BM478282	BM478282 BM478282
c 11	262.8	5.2	271	9	BE762357	BE762357 EM: 1110
12	260.8	5.1	979	11	AK020115	AK020115 Mus musculus
13	259.4	5.1	666	9	BB656292	BB656292 BB656292
c 14	259.6	5.1	408	10	BE543337	BE543337 EST Y
15	256	5.0	581	10	BE520108	BE520108 ; ; 0
16	256	5.0	707	9	BB643458	BB643458 BB643458
17	252.2	5.0	453	9	AW518	AW518 ; ; 0
18	224.8	4.4	596	10	BM088669	BM088669 9e23b04 MA
c 19	214.6	4.2	644	1	BE111214	BE111214 7b44e07.x
20	214.2	4.2	556	1	BE208445	BE208445 qm95f06.y
c 21	207.8	4.1	606	1	BE641699	BE641699 qm96f06.x
22	205.6	4.0	423	9	AA482592	AA482592 zz05h07.r
c 23	197.6	3.9	631	1	BE447130	BE447130 nad3a04.
c 24	185.8	3.6	580	1	BE510655	BE510655 OL H B14
c 25	182.8	3.6	722	12	AZ033629	AZ033629 ZM6197N18
c 26	177.8	3.5	628	9	AI459225	AI459225 tk11b03.s
c 27	175.2	3.5	452	9	AW126794	AW126794 52 M BBL
28	174.2	3.4	479	1	BJ04899	BJ04899 ES1-5b01 BA
29	173.4	3.4	468	1	AZ088573	AZ088573 BE71 25 1
30	169.4	3.3	533	12	AZ088059	AZ088059 BE71 25 1
31	149	2.9	637	9	BE612331	BE612331 BBe12331
32	141.6	2.8	553	10	W47220	W47220 zc39a07.r1
33	140.6	2.8	442	9	BB859598	BB859598 BB859598
34	135	2.7	553	10	BE472859	BE472859 tr92g02.y
c 35	132.8	2.6	534	9	AA972609	AA972609 op95a12.s
c 36	128.2	2.5	258	10	BE848902	BE848902 QV1 EN006
37	121.6	2.4	302	9	AA284278	AA284278 zc39a07.1
c 38	121.2	2.4	289	9	BE162536	BE162536 PM2 B1045
c 39	109.6	2.2	437	10	W47029	W47029 zc39a07.s1
c 40	108.2	2.1	568	9	AI917724	AI917724 tt11c08.x
c 41	107.2	2.1	522	9	AI342006	AI342006 qs98d10.x
42	100.6	2.0	657	12	AZ043795	AZ043795 RP1 23 3
43	98.2	1.9	460	12	AZ486665	AZ486665 IM02BIN20
44	98.2	1.9	607	12	AZ993395	AZ993395 ZM02410F5
c 45	96.4	1.9	449	10	BE058634	BE058634 7k24d09.x

ALIGNMENTS

RESULT 1
BG750735
LOCUS BG750735 606 bp mRNA linear EST 15-MAY-2001
DEFINITION 602706924F1 NIH_MGC_4 Homo sapiens cDNA clone IMAGE:4843636 5', mRNA sequence.
ACCESSION BG750735
VERSION BG750735.1 GI:14061508
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 606)
AUTHORS NIH_MGC_http://www.ncbi.nlm.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, PhD
Email: cmaps-r@mail.nih.gov
Tissue Procurement: A.O.
cDNA Library preparation: Invitrogen Laboratory
cDNA Library Arrayed by: the I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov.
Plate: LLCM1679 row: p column: 05
High quality sequence stop: 606.
FEATURES Location/Qualifiers
source 1..606
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4843636"
/clone_lib="NIH_MGC_4"
/tissue_type="normal pigmented retinal epithelium"
/lab_host="LIBL06 (phage resistant)"
/note="Organ: eye; Vector: pcDB; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT primed, unidirectionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Linda Bond in the laboratory of Gerald M. Rubin (University of

California, Berkeley) using TAG cDNA synthesis kit
(Stratagene) and SuperScript II RT (Life Technologies).
Note: this is a NIH_MGC Library. ["
BASE COUNT 172 a 145 c 144 g 155 t
ORIGIN

Query Match 11.9%; Score 603.4; DB 10; Length 606;
Best Local Similarity 99.8%; Pred. No. 1.3e-149;
Matches 604; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 732 aqaaaatgtaattacctttcctttqqqaaqtccaaqtqtqaqaattacagtqaaaggacc 791
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2 ACAAAATGTAATTCCTCTTCCTTGGGAAGTCCAAGTGTCAGAATTACAGTGAAAGGACC 61

Qy 792 tqc[illegible] 851
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 62 TCCTTA[illegible] 121

Qy 852 [illegible] 911
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 122 CAGCCCCCCCCTTTTTCCTACAAAACACAACACTCCAATTCAGACCCCCTCCCAGAC 181

Qy 912 [illegible] 971
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 182 GCAAACTTTAACATTCCACCACCTCTGATGGCTCATTTCATCTTCAAGACCAGGTACAC 241

Qy 972 [illegible] 1031
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 242 TGCAGCCAAAGACAGCGGPGGTTCAGTTCTTCTTTTACCAGCCCATCAGTCATCAGTGGAG 301

Qy 1032 acaaaactgaattctttccctqcactgtqacgtqtqtqaaqaaattatcaactcaattctqc 1091
||||||||||||||||-|||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 302 ACAAACTGACTTCTTTCCCTGCACTGTGACGTGTGGAGGAGGTTATCAGCTCAATTCTGC 361

Qy 1092 tqaatatqtqaatatcccttqaaqaggqatqattcctaaccattatttqtcactactaccc 1151
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 362 TGAATGTGTGGATATCCCGTTGAAGAGGGTAGTTCCTGACCATTATTGTCACTACTACCC 421

Qy 1152 tqaaaatqtaaaacaaaacaaaactqaaqqaatcaaaatqqatcactqccaatcaaaq 1211
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 422 TGAAAATGTAAAACCAAAAGCAAAACTGAAGGAATGCAGCATGGATCCCTGCCCATCAAG 481

Qy 1212 tqatqqatttaaaqaaqaataatqcctctatqatcaattcaaacctcttcctqctqqaaca 1271
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 482 TGATGGATTTAAAGAGATAATGCCCTATGACCACTTCCAACCTCTTCCTCGCTGGGAACA 541

Qy 1272 taatcctqqaatqaatattccqtqtcctatqqaqqaqqaqattcaqaqacqqaqctttgt 1331
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 542 TAATCCTGGACTGCATTTCCGTGTCCTGTGGAGGAGGGATTCAGAGACGGAGCTTTGT 601

Qy 1332 atqtq 1336
|||||
Db 602 GTGTG 606

RESULT 2
AA707140/c
LOCUS AA707140 485 bp mRNA linear EST 24 DEC 1997
DEFINITION zj49g01.s1 Soares_fetal_liver_spleen_1NFLS_S1 Homo sapiens cDNA
clone IMAGE:452088 3' similar to TR:P97857 P97857 SECRETORY PROTEIN
CONTAINING THROMBOSPONDIN MOTIFS. ;, mRNA sequence.
ACCESSION AA707140
VERSION AA707140.1 GI:2717058
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 485)
AUTHORS Bilter,L., Allen,M., Bowles,L., Dubuque,L., Geisel,G., Jost,S.,
Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin
,J., Moore,B., Scheilenberg,K., Stephe,M., Tan,F., Theisinq,B.,
White,T., Wylie,L., Waterston,R. and Wilson,R.
TITLE Washu-NCI human EST Project
JOURNAL Unpublished (1997)
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 471.
FEATURES Location/Qualifiers
source 1. .485
/organism "Homo sapiens"
/db_xref="GDB:1488444"
/db_xref="taxon:9606"
/clone="IMAGE:452088"
/clone_lib="Soares_fetal_liver_spleen_1NFLS_S1"
/sex="male"
/dev_stage "20 week post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note "Organ, Liver and Spleen. Vector: pT7T3D (Pharmacia)
with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
This is a subtracted version of the original Soares fetal
liver spleen 1NFLS library. 1st strand cDNA was primed
with a Pac I - oligo(dT) primer [5'
AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT 108 a 144 c 114 g 119 t
ORIGIN

Query Match 9.3%; Score 474; DB 9; Length 485;
Best Local Similarity 99.8%; Pred. No. 4.2e-115;
Matches 485; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 236 atqcttqaqqcqactqqaqtqactqctcccaqacctqtqqqqqaqqaqcatcatattctc 295
|||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db 485 ATGCTTGGGGCGACTGGAGTGACTGCTCCCGGACCTGT-GGGGAGGAGCATCATATTCTC 427

Qy 296 tqccqaqatqtqtttqactqqaaqqaattqtqaaaqqccaqaaacattccqtacaaqacatqca 355
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 426 TGCGGAGATGTTTGACTGGAAGCAATTGTGAAGCGCAGAACATTCGGTACAAGACATGCA 367

Qy 356 qcaatcatqactqccctccaqatqcaqaaaatttcaqaqccaqaqtactcaqccctaca 415
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 366 GCAATCATGACTGCCCTCCAGATGCAGAAGATTTCAGAGCCCAGCAGTGCTCAGCCTACA 307

Qy 416 atqtatqtccqtatcaaqqcattatatqaatqqttccacqatataatqatcctqctq 475
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 306 ATGATGTCCACGATATCAGGGGCATTACTATGAATGGCTTCCACGATATAATGATCCTGCTG 247

Qy 476 ccccqtqtqcatcaaqtqtcatqccacaaaqacaaaacttqtatqqtcqaactqcqacctca 535
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 246 CCCCGTGTGCACTCAAGTGTCATGCACAAGGACAAAACTTGGTGGTGGAGCTGGCACCTA 187

Qy 536 aaqtactqqatqqaactcqttqcaacacqqactcctqqacatqtqtatcaqtqqcatct 595
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 186 AGGTACTGGATGGAACTCGTTGCAACACGGACTCCTTGGACATGTGTATCAGTGGCATCT 127

Qy 596 qtcaqqcaqtqqqctqccqatcqqcaactqqqaaqcaatqccaaqqaqqacaactqtqqaq 655
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 126 GTCAGGCAGTGGGCTGCGATCGGCAACTGGGAAGCAATGCCAAGGAGGACAACTGTGGAG 67

Qy 656 tctqtqccqqcaqtqqctccacctqcaqcctqtqtacqaqqacaatcaaaqtcacacqttt 715
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 66 TCTGTGCCGGCGACCTGCAGGCTGTACGGGGACAATCAAAGTCACACGTTT 7

```
qy      716 ctcctq 721
            ||||||
Db        6 ctcctg 1


RESULT  3
AK020961
LOCUS       AK020961                  840 bp    mRNA    linear   HTC 13-JAN-2002
DEFINITION  Mus musculus adult male epididymis cDNA, RIKEN full length
            enriched library, clone:2300119C12:thrombospondin type 1 domain
            containing protein, full insert sequence.
ACCESSION   AK020961
VERSION     AK020961.1  GI:12809613
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (strain:C57BL/6J) adult male epididymis cDNA, mRNA,
            clone lib:RIKEN full length enriched mouse cDNA library
            clone:2300119C12.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High efficiency full length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2  (sites)
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M., Hirozane-Kishikawa,T.,
            Watahiki,A., Gustincich,S., Sandelin,A., Ninomiya,N., Okazaki,Y.
            and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3  (sites)
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishizawa,T., Nagai,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kuochi,K.,
            Sasaoka,T., Inoue,K., Ogawa,Y., Iwaya,M., Watahiki,M.,
            Tamura-Kawakami,K., Yoshino,M., Tanioka,F., Uesugi,S., Hashimoto,J.,
            Sugano,S., Nishikawa,T., Hirakawa,H., Shin-i,T., Kohara,Y.,
            Ishikawa,T., Ogasawara,K., Tanaka,T., Matsubara,K., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20536214
   PUBMED   11076861
REFERENCE   4  (sites)
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5  (bases 1 to 840)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,I., Aono,H., Arai,A.,
            Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Carninci,P.,
            Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hamaguchi,I.,
            Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hirozane-Kishikawa,T.,
            Hume,D., Imotani,K., Ishii,Y., Itoh,M., Iwaya,M., Kaiho,R.,
            Kamiya,H., Kawai,J., Kojima,Y., Konno,H., Kondo,M., Koyama,S.,
            Kurihara,C., Matsuya,Ma,I., Miyazaki,A., Nishi,K., Nomura,K.,
            Numazaki,R., Ohara,M., Okazaki,Y., Okido,T., Owa,T., Quackenbush,J.,
            Saito,R., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D.,
            Schriml,L., Shibata,K., Shibata,Y., Shimasawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tanaka,Y., Taniguchi,S.,
            Tanioka,F., Tejima,Y., Toyo,I., Yamamoto,J., Yamanaka,I.,
            Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and
            Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-AUG-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-jos@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/)
            Fax:81-45-503-9216
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN,
            Division of Experimental Animal Research in Riken, contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer
            [5' GAGAGAGAAAGGAT(AAGAG)3TTTTTTTTTTTTTTVN 3'], cDNA was
            prepared by using trehalose thermo activated reverse transcriptase
            and subsequently enriched for full length by cap trapper. cDNA went
            through one round of normalization to Cot = 10.0 and subtraction to
            Cot = 18%.2. Second strand cDNA was prepared with the primer
            adapter of sequence [5'
            GAGAGAGAATTCTCGAGTTAATTAAATTAATTAA(CCCCCCCCC)3']. cDNA was cleaved
            with BamHI and XhoI. Vector: a modified pBluescript KS(+) after
            bulk excision from Lambda FLC-1. Cloning sites, 5' end: SalI; 3'
            end: BamHI. Host: DH10B.
FEATURES             Location/Qualifiers
     source          1..840
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="MGD:MGI:1911798"
                     /db_xref="taxon:10090"
                     /clone="2300119C12"
                     /sex="male"
                     /tissue_type="epididymis"
                     /clone_lib="RIKEN full length enriched mouse cDNA library"
                     /dev_stage="adult"
     misc_feature    1..840
                     /note="Thrombospondin type 1 domain containing protein
                     data source:Pfam, source key:PF00090, evidence:ISS"
BASE COUNT       227 a    202 c    210 g    199 t
ORIGIN


  Query Match             84.1%;  Score 409.6;  DB 11;  Length 840;
  Best Local Similarity   84.6%;  Pred. No. 8.2e-98;
  Matches  460;  Conservative    0;  Mismatches   84;  Indels    0;  Gaps    0;

Qy      185 accaggcctcaaaaacaatctgttaaatgaaaaaaaaaaatgaaaaaaaaaatgttgtgg 244
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      150 ACAAAGACTCAAAGAAAAGCCAAGCAGAAGAAGACAAAGCAGCCCAAGCAGCCTGGCTGGG 209

Qy      245 gcgaatgaaaatgaaatgaatgccaaaaatgaaggaaaaaaaaaaaaatgaaatgatttgtgg 304
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      210 GCGGAGACTGGAGCCAGTGCTCCCGAGCTGTGCCGAGGAGCCTCTACTCTTTGTGGAGATGG 269

Qy      305 gttttgactgaatgaaaaaatatgtgaaaaaaaaaaaatgtgaatgaaaaaaaaaaatgtgg 364
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      270 GTCTGTACTGGAAGCAAGTGCAAAGGAGCAGAAGATCAGATCAAAGAGCTGCAGCAACAGCG 329

Qy      365 actgcccctccagatgcaaaaatttcaaaccccaaatgctgctcaaaatgaaatgaatgtcc 424
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      330 ACTGCGCCAGCAGATGCAGAAGCAGTCCAGAGCCGACAGCAGTCCAGCTACAATGACGTCC 389

Qy      425 agtatcaaggaaatattactatgaatggctttcaaaatatgaatgatcctgctgcccaggtgt 484
            |||| ||||| ||||| || |  ||| || ||| ||| | |||| ||||||||||| ||||
Db      390 AGTACCAGGTCCAGCAGTACCAGAAGCAGTCCAGCAGTCCAGCTATCAATGACCCTGCTGCC 449

Qy      485 cactcaaatgtcatgaccaaaaaaaaaacttgatgatgaaactgaaaccaacctaaatgactag 544
            |||||||||| || | |||||| | ||||| || |||||| |||||||||| ||| ||| |||
Db      450 CACTCAATGCCATGGCCAAGAAGGTTGGTGATGAGACTGAGACCAACCTGAATGACTAG 509

Qy      545 atgaaactgttgcaaccaatgctgtgaatgatgtgaatgaaaatgaatgtgaaatgaatgaag 604
            |||||||||| ||||| | ||||| ||| | |||| || || || | | | ||| |
Db      510 ATGGAACTGCTGCTGAATGCTGACTCCTCGATATGTGCATAAGCGCAACGCCAGATTG 569

Qy      605 aaatgaaatgaatgaaatgtaatgaatgtcaaaaaaaaaaaatgaatgaatgtgaatgaaag 664
            ||||||| || | || ||||||| |||||||.|||||||| |||| ||||||| |
Db      570 TGGCTGTGATGACAGCTCCTAAGTAATCACAAGGAGCACAATTGCGCGCTGTGCTG 629
```

```
Qy      665 qcgatqqctccacctqcaqqctqtacqqqqqaccaatccaaaqtcacacqttttctcctqaaa 724
            | || ||||||||||||||||| || ||||| || | ||||   ||| | ||||||||| |
Db      630 GGCAGGGTGTGACGTGCAGGTGTGTGGGGGGCACAGTAAGAATGCACCCCTGTCTCACA 689

Qy      725 aaaq 728
            |||
Db      690 AGAG 693


RESULT   4
BF767952
LOCUS       BF767952                 425 bp    mRNA    linear   EST 12-JAN-2001
DEFINITION  CM1 CN0061 201200 672-h01 CN0061 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF767952
VERSION     BF767952.1  GI:12115952
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 425)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brumstein,A., Goo|iveira,R.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202664
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2701422
            Fax: +55-11-2707461
            Email: asimpson@ludwig.org.br
            This sequence was derived from the CAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.lcda.br/scripts_bottom2.pl?CMid=2 CN: CN0061
            201200-672-h01&t4=2000-12-20&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 12
            High quality sequence stop: 425.
FEATURES             Location/Qualifiers
     source          1. .425
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /clone_lib "CN0061"
                     /dev_stage "Adult"
                     /note="Organ: Colon_normal; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into the pOC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT      112 a    100 c     98 g    115 t
ORIGIN


  Query Match             7.9%;  Score 401.4;  DB 10;  Length 425;
  Best Local Similarity   99.5%;  Pred. No. 9.4e-96;
  Matches  414;  Conservative    0;  Mismatches    1;  Indels    1;  Gaps    1;

Qy      915 aacttttaaqattccaqqacctctqatqqctqatttcatcttcaaqaccaqqtacactqc 974
            ||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db       12 AACTTTTAAGATTCCAGGACCTCTGATGGC-GATTTCATCTTCAAGACCAGGTACACTGC 70

Qy      975 aqccaaqacaqcatqttcaqttcttctttttaccaqccatcaqtcqtcaqtqqaqaca 1034
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       71 AGCCAAAGACAGCGTGGTCAGTTCTTCTTTTATCAGCCCATCAGTCATCAGTGGAGACA 130

Qy     1035 aaatqaqttttttbbbbtaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1094
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      131 AACTGACTTCTTTGCCTGCACTGTGACGTGTGGAGGAGGTTATCAGCTCAATTCTGCTGA 190

Qy     1095 atqtqtggatatccqcttqaaqaqqqtaqttcctqaccattattqtcactactaccctqa 1154
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      191 ATGTGTGGATATCCGCTTGAAGAGGGTAGTTCCTGACCATTATTGTCACTACTACCCTGA 250

Qy     1155 aaatqtaaaaaaaaaaaaaaaaaaaqaqqaaatqaqcataaaaaaaaaaaaaaaaaaaaa 1214
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      251 AAATGTAAAACTAAAACTAAAATTGAAGAAATGTAGTATTAATTTCCTCCCACTAAGTGA 310

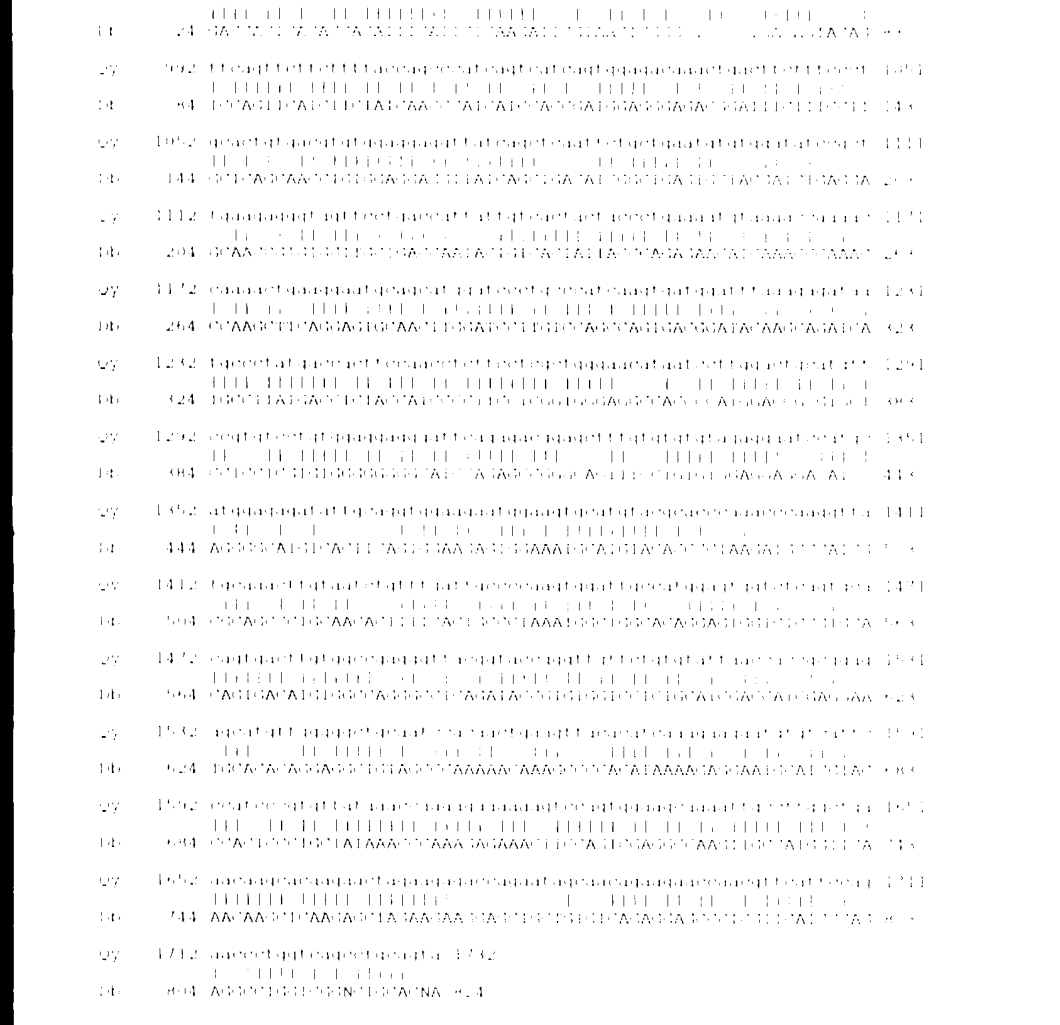
Qy     1215 tqaattttaaaqaqataatqccctatqaccacttccaaccctcttcctcqctqqqaacataa 1274
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      311 TGGATTTAAAGAGATAATGCCCTATGACCACTTCCAACCCTCTTCCTCGCTGGGAACATAA 370

Qy     1275 tcctqqactqcatqttccqtgtcctqtqaqaaaaqqaattcaaaaqcqaaqttt 1329
            |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      371 TCCTTGGACTGCATGTTCCGTGTCCTGTGAGGAGGGATTCAGAGACGGAGCTTT 425


RESULT   5
BM476141
LOCUS       BM476141                 1009 bp    mRNA    linear   EST 05-FEB-2002
DEFINITION  AGENCOURT_6478885 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:5558669
            5', mRNA sequence.
ACCESSION   BM476141
VERSION     BM476141.1  GI:18525183
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1009)
  AUTHORS   NIH MGC Project [Web] [ ] [ ] [ ].
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: the I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12282  row: a  column: 06
            High quality sequence stop: 647.
FEATURES             Location/Qualifiers
     source          1. .1009
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /clone "IMAGE:5558669"
                     /clone_lib="NIH_MGC_88"
                     /tissue_type "duodenal adenocarcinoma, cell line"
                     /lab_host "DH10B (phage-resistant)"
                     /note "Organ: small intestine; Vector: pCMV SPORT6;
                     Site_1: Not1; Site 2: Sal1; Cloned unidirectionally;
                     oligo-dT primed. Average insert size 1.767 kb. Library
                     enriched for full-length clones and constructed by Life
                     Technologies. Note: this is a NIH MGC Library."
BASE COUNT      234 a    299 c    268 g    204 t      4 others
ORIGIN

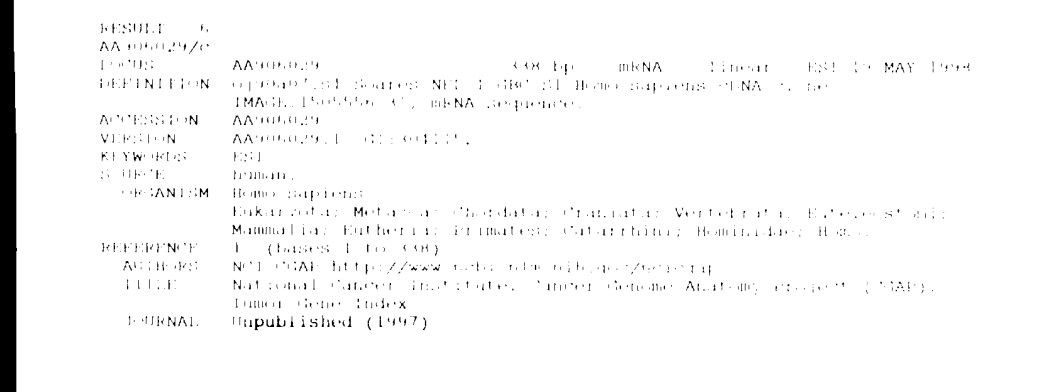

  Query Match             7.8%;  Score 394.2;  DB 10;  Length 1009;
  Best Local Similarity   68.2%;  Pred. No. 1.1e-93;
  Matches  546;  Conservative    0;  Mismatches  255;  Indels    0;  Gaps    0;

Qy      932 qacctctqatqqctqatttcatcttcaaqaccaqqtacactqcaqccaaqacaqcatqtqq 991
```



RESULT 6
AA006029/c
ID AA006029 standard; mRNA 438 BP. 11-MAY-1998
DEFINITION yo04a07.r1 Soares NFL T1GBC Homo sapiens cDNA 5', mRNA sequence.
ACCESSION AA006029
VERSION AA006029.1 GI:1464135
KEYWORDS EST.
SOURCE human.
  ORGANISM Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Ho...
REFERENCE 1 (bases 1 to 438)
  AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap
  TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
  Tumor Gene Index
  JOURNAL Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.
  Email: cgaps@mail.nih.gov
  This clone is available royalty-free through LLNL; contact the
  IMAGE Consortium (info@image.llnl.gov) for further information.
  Insert length: 508 Std Error: 0.00
  Seq primer: 43mer Rxn. Kit from Amersham
  High-quality sequence stop: 278.
FEATURES Location/Qualifiers
  source 1..438
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /clone="IMAGE:1565956"
  /clone_lib="Soares_NFL_T1GBC_S1"
  /tissue_type="HBL96"
  /note="Organ: pooled; Vector: pT7T3-Pac (Pharmacia) with
  a modified polylinker; Site 1: Not I; Site 2: Eco RI;
  Equal amounts of plasmid DNA from three normalized
  libraries (fetal lung NbHL19W, testis NbHT, and B cell
  NCI_CGAP_GCB1) were mixed, and ss circles were made in
  vitro. Following BAP purification, this DNA was used as
  tracer in a subtractive hybridization reaction. The driver
  was PCR-amplified cDNAs from pools of 5,000 clones made
  from the same 3 libraries. The pools consisted of
  I.M.A.G.E. clones 294180-302087, 082452-087259,
  726408-728711, and 729096-731595. Subtraction by Bento
  Soares and M. Fatima Bonaldo. "
BASE COUNT 73 a 68 c 75 g 121 t
ORIGIN



RESULT 7
BB643755
ID BB643755 standard; mRNA 622 BP. 26-OCT-2000
DEFINITION BB643755 RIKEN full-length enriched, adult male spinal cord Mus musculus cDNA clone A430072P05, mRNA sequence.
ACCESSION BB643755
VERSION BB643755.1 GI:19471129
KEYWORDS EST.
SOURCE house mouse.
  ORGANISM Mus musculus
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 622)
  AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hamaguchi,I., Hara,A.
  , Hirozane-Kishikawa,T., Ishii,Y., Itoh,M., Kawai,J., Konno,H., Kouda
  ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
  Okazaki,Y., Ooe-Ishi,A., Saito,R., Sakai,C., Sakazume,N., Satomi,S., Shibata,K., Shinagawa,A., Shiraki,T., Soma-Nagae,Y., Suzuki,H.,

```
            Taqami,M., Taqawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81 45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://[illegible]
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of Cap-trapper selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. , 10 (10), 1617-1630 (2000)
             Waqi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,I., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res. ,
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. , 11 (2), 291-299 (2001)
             Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
            ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
            Hayashizaki,Y.
             Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
             Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
             6 mouse tissues.
FEATURES             Location/Qualifiers
     source          1..622
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="A330077F03"
                     /clone_lib="RIKEN full length enriched, adult male spinal
                     cord"
                     /sex="male"
                     /tissue_type="spinal cord"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCGAAGACGTCTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     C0 Rot = 10.0 and subtraction to Rot = 459.0. Second
                     strand cDNA was prepared with the primer adapter of
                     sequence [5' GAGAGAGAGATCTCGAGTTAATTAAAATTAATCCCCCCCCCCCCCC
                     3']. cDNA was cleaved with XhoI and BamHI. Vector: a
                     modified pBluescript KS(-) after bulk excision from Lambda
                     FLC 1. "
BASE COUNT       151 a    157 c    172 g    142 t
ORIGIN


  Query Match             5.8%;  Score 293.6;  DB 9;  Length 622;
  Best Local Similarity   69.1%;  Pred. No. 6.5e-67;
  Matches  401;  Conservative    0;  Mismatches  179;  Indels    0;  Gaps    0;
```

```
Qy      186 acagaaatcaaqqaaaaaatcagttaagatgaaaaaaaaaatggaaaatqqaaatcttqaaaa 245
            || ||  || || |   | || ||||| ||||||   || |||  |||||||| |||||
Db       25 CCTGAGTTCCAGGACTGCACGCTCAGAGGAAGACCGTGAAGGCCTCTGGGATGCCTGGGG 84

Qy      246 cqaatqaaatqaatgctcctcaaaaatqtqqqqqaqqqaatcatatttctqtqqaqaatq 305
            |   ||||| || || ||| | |||||||| || || || || ||||| ||| |   | ||
Db       85 CCCCTGGAGCGAGTGTTCCAGAACCTGTCGTGGGGTCTCCTGTATTCCCTGAGACGCTG 144

Qy      306 tttqactqqaaaqqaattqtqaaaqqqcaqaaacatttcqqttacaaaqaatqtcaqcaatcatqa 365
            |||   | | ||  |||||||||   || ||||| || |  |||||||| |||   ||
Db      145 CCTGAGCAGCAAGAGCTGTGAAGGGAGAAATATTCGATATAGAACATGCAGTAATGTGGA 204

Qy      366 ctqcctctaaatatcaqqqqqattcaatqccaqcaatqatqctcaqccttaccaatqatqtqtca 425
            |||||| ||||| |||| |||||| |||| || |||| || ||   | || || ||| |
Db      205 CTGCCCACCAGAAGCAGTGATTTCCGAGCTCAACAGTGTTCTGCTCATAACGACGTCAA 264

Qy      426 qtatcaqqqqcattacttatqaatqqcttcaqcqatatqaatqatcctqctqccctqtqtq 485
            ||| || || ||    |||||||||||||||   || |||||| || |     || || |
Db      265 GTACCATCGACAGCTTTATCAATGGCTTCCTCGTATCTAATGAGTCCGACAATCTATGCTT 324

Qy      486 actcaqqtqtcatqcaccaqqqcaaaaattqatqatqtqqqctqqcaccttaaqqtacttqqa 545
            ||||||||| || ||| ||||    | | |||| |||| || |||| ||||    | ||
Db      325 ACTCAACTTCCCAAGCAAAACCAACTCAGCTTCTTTCTCAACTAGCACCGAAGGTCTTATCA 384

Qy      546 tqqaactcqttqcaaacaqqaactcctttqqacatqtqtatcaqtqqcatctqtcaqqcaqt 605
            ||| || || ||| | || || ||  || | |||| |||| || | || ||    ||
Db      385 TGGTACGCGATGCTATACAGAATCACTGGATATGTGCATCAGCGGCCTATGCCAAATTGT 444

Qy      606 gqqctqtqatqctqqaqqcaqtcctqtqqqaqtctqtqccaq 665
            |||| |||   || |||| |||| || ||| |||| ||||||||| ||   |||
Db      445 TGGCTGTGATCACCAGCTGGGGAGCACTGTCAAAGAGGATAACTGTGGAGTGTGCCACGG 504

Qy      666 cqatqqctcaqcctqcaqqcttqtacqqqqcaatcaaaqtcacaqtttctcctqaaaa 725
            |||||||| || ||| |||| || || || |  || || ||  | ||| |    ||
Db      505 AGATGGCTAAGGTGCCCCTCCCCCGAACCCACTATAAATCTCAGCTGTCTGCAAGTAA 564

Qy      726 aaqaqaaqaaaaatqtaattqctqttcctttqqaaaqtcaa 765
            |   || || |  ||  | ||  ||||||  || ||  ||
Db      565 ATCGGATGACACCGTGGTCGGCCATTGCTTACGGCAGCAGA 604
```

```
RESULT   8
AQ743958
LOCUS       AQ743958                 890 bp    DNA     linear   GSS 16-JUL-1999
DEFINITION  HS_5504_A1_E01_17A RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate-1080 Col-1 Row-G, DNA sequence.
ACCESSION   AQ743958
VERSION     AQ743958.1  GI:5521480
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 890)
  AUTHORS   Muhairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,I.,
            Keller,A., Chabot,E., Fulton,J., Yonas,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Muhairas GG. Wallace JC. Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-4618
            Fax: (206) 616-4887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://[illegible]/ordering_bac.htm)
```

```
            of from Research Genetics (info@resgen.com). BAC end Web Server
            http://www.htsc.washington.edu
            Plate: 1080 row: [illegible] column: 1
            Seq primer: [illegible]
            Class: BAC ends
            High quality sequence stop: [illegible].
FEATURES             Location/Qualifiers
     source          1..[illegible]
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate 1080 Col 1 Row [illegible]"
                     /clone_lib="RPCI-11 Human Male BAC Library"
                     /sex="male"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     Male blood DNA was isolated from one randomly chosen donor
                     and partially digested with a combination of EcoRI and
                     EcoRI Methylase. Size selected DNA was cloned into the
                     pBACe3.6 vector at EcoRI sites"
BASE COUNT      228 a    [illegible] c    218 g    2[illegible] t
ORIGIN

  Query Match             [illegible]%;  Score 281.[illegible];  DB [illegible];  Length [illegible];
  Best Local Similarity   [illegible]%;  Pred. No. [illegible];
  Matches  286;  Conservative    [illegible];  Mismatches   [illegible];  Indels   [illegible];  Gaps   [illegible];

Qy       1696 [illegible] 1755
Db        125 [illegible] 184

Qy       1756 [illegible] 1815
Db        185 [illegible] 244

Qy       1816 [illegible] 1875
Db        245 [illegible] 304

Qy       1876 [illegible] 1935
Db        305 [illegible] 364

Qy       1936 [illegible]
Db        365 [illegible] 41[illegible]


RESULT 9
BE888902
LOCUS       BE888902         850 bp    mRNA    linear   EST [illegible]
DEFINITION  601[illegible]L88281 NIH_MGC_[illegible] Homo sapiens cDNA clone IMAGE:[illegible]
            mRNA sequence.
ACCESSION   BE888902
VERSION     BE888902.1  GI:[illegible]
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to [illegible])
  AUTHORS   NIH MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM[illegible] row: q column: [illegible]
            High quality sequence stop: 742.
FEATURES             Location/Qualifiers
     source          1..850
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3915294"
                     /clone_lib="NIH_MGC_71"
                     /tissue_type="leiomyosarcoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 2.1 kb. "
BASE COUNT      202 a    219 c    228 g    201 t
ORIGIN

  Query Match             5.[illegible]%;  Score 278;  DB [illegible];  Length 850;
  Best Local Similarity   67.[illegible]%;  Pred. No. 1.1e-62;
  Matches  451;  Conservative    [illegible];  Mismatches  21[illegible];  Indels    4;  Gaps    4;

Qy        186 [illegible] 245
Db        103 [illegible] 162

Qy        246 [illegible] 305
Db        163 [illegible] 222

Qy        306 [illegible] 365
Db        223 [illegible] 282

Qy        366 [illegible] 425
Db        283 [illegible] 342

Qy        426 [illegible] 485
Db        343 [illegible] 402

Qy        486 [illegible] 545
Db        403 [illegible] 462

Qy        546 [illegible] 605
Db        463 [illegible] 522

Qy        606 [illegible] 665
Db        523 [illegible] 581

Qy        666 [illegible] 725
Db        582 [illegible] 641

Qy        726 [illegible] 782
Db        642 [illegible] 701

Qy        783 [illegible] 842
Db        702 [illegible] 761

Qy        843 [illegible] 852
Db        762 [illegible] 771


RESULT 10
BB478282
LOCUS       BB478282         6[illegible] bp    mRNA    linear   EST 25-OCT-2000
DEFINITION  BB478282 RIKEN full-length enriched, 13 days embryo heart Mus
```

```
            musculus cDNA clone D330029102 3', mRNA sequence.
ACCESSION   BB478282
VERSION     BB478282.2  GI:16418420
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 619)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.
            , Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
            ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
            Okazaki,Y., Shibata,K., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
            ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
            Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
  JOURNAL   Unpublished (2001)
COMMENT     On Jul 23, 2000 this sequence version replaced gi:9395891.
            Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://[illegible]
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
             [illegible],K., Fujiwake,?., Inoue,K., Togawa,Y., Izawa,M., Hara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
             Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
            ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
            Hayashizaki,Y.
             Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
             Please visit our web site (http://[illegible]) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
FEATURES             Location/Qualifiers
     source          1..619
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="D330029102"
                     /clone_lib="RIKEN full-length enriched, 13 days embryo
                     heart"
                     /tissue_type="heart"
                     /dev_stage="13 days embryo"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAGAGGGCGCAACTCGAGTTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. Second strand cDNA was prepared with the
                     primer adapter of sequence [5'
                     GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCC 3']. cDNA
                     was cleaved with BamHI and XhoI. Vector: a modified
                     pBluescript KS(+) after both excision from Lambda FLC I."
BASE COUNT      160 a    142 c    169 g    148 t
ORIGIN

  Query Match             5.1%;  Score 274;  DB 9;  Length 619;
  Best Local Similarity   70.8%;  Pred. No. 1.1e-61;
  Matches  364;  Conservative    0;  Mismatches  150;  Indels    0;  Gaps    0;

Qy        215 aagacaaagatggcaactgggatgcttggagccgactgggatgactgctcccagaacctgtg 274
Db          1 AAGACCGTGAAGGCTTCTGGGATGCTTGGGGCCCTTGGAGCCGAGTGTTCCAGAACCTGTG 60

Qy        275 ggggaggagcatcatattctctgcggagatgtttgactggaaggaattgtgaagggcaga 334
Db         61 CTGCCGTCCTCTTATTCCCTGACACCTTGCTTCACTACCAACACCTCTGAAGGGACAA 120

Qy        335 acattcggtacaagacatgcagaatcatgactgcctccagatgcagaagatttcagag 394
Db        121 ATATTCCATATAGAACATGCAGTAATGTGACTGCCCACCAGAAGCAGGTGATTTCCGAG 180

Qy        395 ccaagcagtgctcagcctacaatgatgtccagtatcagggcattactatgaatggcttc 454
Db        181 CTCAACAGTGTTCTGCTCATAACGACCTCAAGTACCATGGACAGCTTTATGAATGGCTTC 240

Qy        455 [illegible] 514
Db        241 CTGTATTTAATCACCCGACAATCCATCTTCATTCAAGTCCTAAGTAAAACGAACCAGCC 300

Qy        515 tggtggtggaagctggcacctaagatactggatagaactcattgcaacacgaactccttgg 574
Db        301 TGGTTGTGGAACTAGCACCGAAGGTCTTAGATGGTACGCCATGCTATACAGAATCACTGG 360

Qy        575 acatgtgtatcagtggcatctgtcaggcaatggagctgcgatcggcaactgggaagcaatg 634
Db        361 ATATGTGCATCAGGGGCCTATGCCAAATTGTTGGCTGTGATCACCAGCTGGGGAGCACTG 420

Qy        635 [illegible] 694
Db        421 TCAAAGAGGATAACTGTGGAGTGTGCAACGGAGATGGCTCAACGTGCCGGCTGGTCCGAG 480

Qy        695 gacaatcaaagtcacacgtttctcctgaaaaaag 728
Db        481 GGCAGTATAAATCTCAGCTCTCTGCAAGTAAATG 514


RESULT 11
BE162337/c
LOCUS       BE162337                 271 bp    mRNA    linear   EST 21-JUN-2000
DEFINITION  PM2-HT0451-080100-003-E04 HT0451 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BE162337
VERSION     BE162337.1  GI:8625058
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 271)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
```

```
                Sequence Lens
  JOURNAL       Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE       20219064
  COMMENT       Contact: Simpson A.J.G.
                Laboratory of Cancer Genetics
                Ludwig Institute for Cancer Research
                Rua Prof. Antonio Prudente 109, 4 andar, 01509-010 - Sao Paulo - SP
                Brazil
                Tel: +55 11 3274 720
                Fax: +55 11 3767 4
                Email: asimpson@ludwig.org.br
                This sequence was derived from the FAPESP/LICR Human Cancer Genome
                Project. This entry can be seen in the following URL:
                (http://www.ludwig.org.br/cgi-bin/db_entry.pl?IMAGE=3491491 or
                http://www.ludwig.org.br/cgi-bin/db_entry.pl?IMAGE=3491491)
                Seq primer: pUC 18 forward
                High quality sequence start: 1
                High quality sequence stop: 271.
FEATURES                Location/Qualifiers
     source             1..271
                        /organism="Mus sapiens"
                        /db_xref="taxon:9606"
                        /clone_lib="unknown"
                        /dev_stage="Adult"
                        /note="Organ: head neck. Vector: pUC18; Site 1: BamH
                        Site 2: SmaI; A mini library was made by cloning products
                        derived from a REEST PCR (U.S. Patent application
                        No. 09/...... Ludwig Institute for Cancer Research)
                        profiles into the pUC 18 vector. Reverse transcription of
                        tissue mRNA and cDNA amplification were performed under
                        low stringency conditions."
BASE COUNT         69 a      65 c      58 g      80 t
ORIGIN


  Query Match             5.8%;  Score 267.8;  DB 9;  Length 271;
  Best Local Similarity   96.8%;  Pred. No. 3.5e-05;
  Matches  262;  Conservative    0;  Mismatches    2;  Indels    1;  Gaps    1;

Qy     740 qaaaaaaatqtaattqtqttcctttaqaaaatcaaatqtqaaqatttcaatcaaaaaa 799
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     271 GAAGAAAATGTAATTGCTGTTCCTttcAGAAAtCAAATGTGAGAATTCAATCAAAAAA 212

Qy     800 cctqcccqcctctttattqcatqcatcaaaaacttcaaaqaqcaaaaaaaaaaaaaaaa 859
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     211 CCTGCCCACCTCTTTATTGCATGCATCAAAAACTTCAAAGAGCAAAAAAAAAAAAAAAA 152

Qy     860 aaaaaaaaqqqqqatttttqtttqtqtaqaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 919
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     151 AACAAGGGGGGGGCTTTTGTTGTGTAGAAAAACTACAACTAGTGGAATTTCAAGAGGGGTTGTGAA 92

Qy     920 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 979
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      91 AAGCAAACTTTAAGATTCTACTGTACTGTGATAGCTCATTTCATCTTCAAAACACTGTTA 32

Qy     980 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1039
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      31 ACTTCTACATCAAAGACACTCTGAATCATCTTCT 1

RESULT 12
AK020115
LOCUS       AK020115                 970 bp    mRNA    linear   HTC 16-JAN-2002
DEFINITION  Mus musculus 12 days embryo male wolffian duct includes surrounding
            region cDNA, RIKEN full-length enriched library,
            clone:2700020A06 product:unknown type I domain containing protein,
            full insert sequence
ACCESSION   AK020115
VERSION     AK020115.1  GI:12963655
KEYWORDS    HTC; cDNA trapper.
SOURCE      Mus musculus (strain:C57BL/6J) 12 days embryo male wolffian duct
            includes surrounding region cDNA to mRNA, clone:RIKEN
            full-length enriched mouse cDNA library clone:2700020A06 product:
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
  REFERENCE 1  (sites)
    AUTHORS Carninci,P. and Hayashizaki,Y.
    TITLE   High efficiency full-length cDNA cloning
    JOURNAL Meth. Enzymol. (in press)
    MEDLINE 99279253
     PUBMED 10449636
  REFERENCE 2  (sites)
    AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
    TITLE   Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
    JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
    MEDLINE 20499374
     PUBMED 11042159
  REFERENCE 3  (sites)
    AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,O.,
            Yoneda,Y., Ishikawa,I., Ozawa,K., Tanaka,I., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
    TITLE   RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
    JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
    MEDLINE 20560104
     PUBMED 11076861
  REFERENCE 4  (sites)
    AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
    TITLE   Functional annotation of a full-length mouse cDNA collection
    JOURNAL Nature 409, 685-690 (2001)
  REFERENCE 5  (bases 1 to 979)
    AUTHORS Adachi,J., Aizawa,K., Akahira,S., Akimura,I., Arai,G., Arai,A.,
            Arakawa,T., Ballarotti,K., Bono,H., Browstein,M., Bult,C.,
            Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hamaraki,I.,
            Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hiraoka,I., Hori,F.,
            Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kashiwagi,I.,
            Kato,S., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S.,
            Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K.,
            Numazaki,K., Ohno,M., Okazaki,Y., Okido,I., Owa,C., Qureshenbush,J.,
            Saito,H., Saito,K., Sakai,C., Sakai,K., Sano H., Sasaki,D.,
            Schriml,L., Shibata,K., Shibata,Y., Shimaqawa,A., Shiraki,I.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,I., Tejima,Y., Teya,I., Yamamoto,I., Yamanaka,I.,
            Yasunishi,A., Yoshida,S., Yoshino,M., Muramatsu,M. and
            Hayashizaki,Y.
    TITLE   Direct Submission
    JOURNAL Submitted (08-AUG-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/) Tel:81-45-503-9210
            Fax:81-45-503-9216)
  COMMENT   Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN,
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer
            5' GAGAGAGAAGAGATC'AAGAG'3'(TTTTTTTTTTTTTTTTVN 3'), cDNA was
            prepared by using trehalose thermo activated reverse transcriptase
            and subsequently enriched for full length by cap trapper. Second
            strand cDNA was prepared with the primer adapter of sequence 5'
            GAGAGAGAAGAGATCAAGAGTCTTTTTTTTTTTTTTTTVN 3'. cDNA was cleaved
            with BamHI and XhoI. Vector: a modified pBluescript SK(+) after
            bulk excision from Lambda ZAP II. Cloning sites: 5' end: SalI; 3'
            end: BamHI, NotI, XhoI.
```

```
FEATURES             Location/Qualifiers
     source          1..979
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="MGI:MGI:1913602"
                     /db_xref="taxon:10090"
                     /clone="6720426B09"
                     /sex="male"
                     /tissue_type="urothelium dome (includes surrounding region)"
                     /clone_lib="RIKEN full length enriched mouse cDNA library"
                     /dev_stage="12 days embryo"
     CDS             85..663
                     /codon_start=1
                     /product="BAP22000.1"
                     /db_xref="GI:13967706"
                     /translation="MECCPRAAPCTPLLVLAFLLLSSRTARSEEDRECLWDAWCPWSE
                     CSRTCGGCAYYCRPVLSSPSPCPSNIPFFTSNVCRTAAAPPAAPPVYRH
                     CLEFWNLFVSNLLSNCSSASPAPCLAVFLARSALSVCFLSLFPISLPLV
                     GCDHQLSSIVRRNCCVCNCCSSCPVCPRAV"
BASE COUNT       252 a    226 c    269 g    232 t
ORIGIN


  Query Match          5.3%;  Score 266.8;  DB 11;  Length 979;
  Best Local Similarity   70.2%;  Pred. No. 1.1e-56;
  Matches  358;  Conservative    0;  Mismatches  152;  Indels    0;  Gaps    0;

Qy     186 ccaagacctcaagaaaaactcgttcagatgaaagcaaaatatgcgaactgcaaatgcttcgaa 245
           || ||  || || |   | || ||||| ||||||   || |||  |||||||||| |||||
Db     144 CCTGAGTTCCAGGACTCCACGCTCAGACGAAGACTGCTSAAGCTTTTTGGCAATCCCTCCCT 203

Qy     246 cgactgaagtgactgtgtcccccctgtgtgggggggtgcatatttgtctgcaaaatg 305
           |   ||||| || || ||| | |||||||| || || || || ||||| ||| |   | ||
Db     204 CCCTTGGAGCGAGTTCCAGAACCTCGTCCTCGCGGCTCCCTCCTATTCCCTGAGACCCTC 263

Qy     306 tttgactgaaaaaaatttgtgaaaggcagaaacattcgatacaaaacatgcaacaaatcatga 365
             |||   | | ||  ||||||||   || ||||| || |  |||||||| |||   ||
Db     264 CCTGAGCAGCAACAGCTGTGAACGGAGAAATATTCGATATAGAACATGCAGTAATGTGGA 323

Qy     366 ctgccctccagatgcaaaaaatttcaaaaccccaaccaatcttcaaactacaatgaatgtcca 425
           |||||| ||||| |||  |||||| |||| || ||||| || ||  | || || ||| |
Db     324 CTGCCCACCACAAGCAGCTGATTTCCCAGCTCAACAGTGTTCTGCTCATAACGACGTCAA 383

Qy     426 gtatcaaaaaacattactatgaatgtgcttcaagataatgtgatctgtcttgtgaaatgtatca 485
           ||| || || ||  ||||||||||||||   || |||||| || |     || ||  |
Db     384 GTACCATGGACAGCTTTATGAAIGGCTTCCTGTATCTAATGACCCCGACAATCCATGCTC 443

Qy     486 actcaaatgtcatgtcacaaaqcaacaaaattttgatagtgaaactgcaacctaaaatactgga 545
           |||||||||| || ||| ||||   | | |||| |||| || |||| |||||  |  |
Db     444 ACTCAAGTGCCAAGCAAAAGGAACCAGCCTGGTTGTGGAACTAGCACCGAAGGTCTTATA 503

Qy     546 tggaatctgttgtaaacatttatcttgaacatgtgtatcatcgtgaatatgtaggaatat 605
           ||| || || ||| | || || |  |||| |||| ||||| ||| | || ||    ||
Db     504 TGGTACGCGATGCTATACAGAATCACTGGATATGTGCATCACGGGCCTATGCCAAATTGT 563

Qy     606 qqqctqcqatcqtcqcaactcqqaaatcaaatcaaaaaaatcqtqqaatctatqtqcaa 665
           |||| ||||  || | | |||| || ||| |||  |||||||||| ||   |||
Db     564 TGGCTGTGATCACAGCTCGGGAGCACTGTCAAAAAGGATAACTGTGGAGTGTGCAACGG 623

Qy     666 cqatqqctcaaatcqaqatttqtaaaat 693
           ||||||| || ||| ||  | | |
Db     624 AGATGGCTCAACGTGCGGGTTGGTCCGAG 653


RESULT  13
BB659292
LOCUS       BB659292                 660 bp    mRNA    linear   EST 26 OCT 2001
DEFINITION  BB659292 RIKEN full length enriched, 12 days embryo heart Mus
            musculus cDNA clone D330029I02 5', mRNA sequence.
ACCESSION   BB659292
VERSION     BB659292.1  GI:16493113
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 660)
  AUTHORS   Arakawa,T., Carninci,P., Shibata,K., Yamanaka,R., Muramatsu,M.,
            Hara,A., Hirozane-Kishikawa,T., Hirokawa,N., Konno,H., Kondo,S.,
            Hayashizaki,Y., Watahiki,M., Kawai,J., Shibata,K., Yoshiki,A.,
            Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,K.,
            Sasaki,D., Shiraki,T., Kawaji,H., Shibata,Y., Sugiyama,D.,
            Tagami,M., Tagawa,A., Takahashi,F., Takeda,S., Tsuchiya,S.,
            Watanabe,S., Yasuda,J., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse FSIs (Arakawa,T., et al. 2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81 45 503 9222
            Fax: 81 45 503 9216
            Email: genome-res@gsc.riken.go.jp
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Normalization and subtraction of cap-trapper selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. 10 (10), 1617-1630 (2000)
            Waga,I., Fujiwake,S., Inoue,K., Tegawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,T., Hira,A. and
            Hayashizaki,Y.
            RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
            Kondo,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
            Computer-based methods for the mouse full length cDNA
            encyclopedia: real time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
            Kondo,S., Shimogawa,A., Saito,T., Fitzsawa,K., Yamanaka,I., Aizawa
            ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
            Hayashizaki,Y.
            Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences. Mamm. Genome, 12, 673-677 (2001)
            Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
            e mouse tissues.
FEATURES             Location/Qualifiers
     source          1..660
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="D330029I02"
                     /clone_lib="RIKEN full-length enriched, 12 days embryo
                     heart"
                     /tissue_type="heart"
                     /dev_stage="12 days embryo"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN, Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAGCGGCCGCAATCGATTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermoactivated reverse
                     transcriptase and subsequently enriched for full length by
                     cap trapper. Second strand cDNA was prepared with the
                     primer adapter of sequence [5'
                     GAGAGAGAGATTCTCGAGTTAATTAAATTAATCGCCGCCGCCGCCC 3']. cDNA
```

was cleaved with BamHI and XhoI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda ZAP II."
BASE COUNT 161 a 163 c 141 g 145 t
ORIGIN

Query Match 5.2%; Score 202.4; DB 9; Length ??;
Best Local Similarity 20.6%; Pred. No. 1.4e-58;
Matches 65; Conservative 0; Mismatches 145; Indels 32; Gaps 0;

Qy 186 ccaaactcaaaaaaactcgtt caaatcaaacaaaatgaaaactgaaatgttaaa 245
Db 164 CCTGAACCCCAAGAACTGCAACGCCTAAGAGGAAGACCGTCGAAAGGCCTCTGGACGCCTCCCC 223

Qy 246 cgactgaatgactgctccgaaatctgtgggggggaaatgatattctctgaaaatga 305
Db 224 CCCTCGAGCCCACTGTTCCCACGAACCTGTGGCTGCGGGGTGCCTCTATTCCTGAACGCGTCT 283

Qy 306 tttgactgggaaggaattgtgaaaaaacgaattgtaaaaaacatgaacaatgtgtta 365
Db 284 CCTGAGCACGTAAGAGCTGTCGAACGCGAGAAATATTCCGATATACAACATCCACTAATCTCTA 343

Qy 366 ctgccctcaaatgcgaaatcttcaaggcaagcaaatgctccgccaaaatatatcaa 425
Db 344 CTGCCCACCAGAAGCAGCTGATTCGCAGCTCAACATCGTCTGCTCATAACGACTCTAA 403

Qy 426 gatccgggcgattactatgaatgattccgaaatgaatatcctgacgccgttgtgatc 485
Db 404 CTACCATCGCACACCTTCTACTAATCCGTTGTCTGAACTACATCCGCCAGCAACTCTTCTATA 463

Qy 486 aatcaatgtgtatgcaaccgaaaccttcttgaatgaaatcaatacaaaccatatcaa 545
Db 464 ACTCAACGTCCCAACGAAAACGAACGCAGTCCGTTGCTGAACTACATCAGCCAGCTTCTATA 523

Qy 546 tgaaactgttgtcgcaacgaaccgcttcgttgaatgtgatgaatgtaatgaaatta 605
Db 524 TGCTACCGCAGTGCTATACACGAACTACTGCATATCGTCCATCAGCCGCTATCTAAAATCC 583

Qy 606 ggcctgcatcgcgaatcgtgatgcgaatgaaatcgcgtatgtatttcgcgaa 665
Db 584 TGCCTCCGAACACTACTACGTTCCCACTACGTTCAAAGAAGATAACTAACTCCGAACTCCAACTC 643

Qy 666 cgatgctcccgctgc 681
Db 644 AGATCGCTCAACTCGC 659

RESULT 14
BF543337/c
LOCUS BF543337 408 bp mRNA linear EST 13-DEC-2000
DEFINITION MR_K_Y0_aba_a_03_r_01_c1_f1_k_Yo Rattus norvegicus cDNA clone MR_K_Y0_aba_a_03_r_01_c1, mRNA sequence.
ACCESSION BF543337
VERSION BF543337.1 GI:11614444
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 408)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044177
COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319-335-8750
Fax: 319-335-9006
Email: mbsoares@blue.weeg.uiowa.edu
cDNA library preparation: MRK, Soares Lab's one distribution: clones will be available through Research Genetics (www.resgen.com) this clone is also available through the I.M.A.G.E. Consortium at LLNL (info@image.llnl.gov). IMAGE ID: 1790644
Seq primer: M13 Forward.
FEATURES Location/Qualifiers
source 1..408
/organism="Rattus norvegicus"
/strain="Sprague Dawley"
/db_xref="taxon:10116"
/clone="MR_K_Y0_aba_a_03_r_01"
/clone_lib="MR_K_Y0"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT7T3-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The MR_K_Y0 library is a subtracted library derived from an individually tagged normalized whole eye (minus the lens) library. The driver for the subtraction consisted of a pool of all previous libraries (MR_K_A0, MR_K_A1, MR_K_E0, MR_K_E1, MR_K_C0, and MR_K_C1). The tag is a string of 5'5 nucleotides present between the Not I site and the oligo dT track which allows identification of the library of origin of a clone within the mixtures. The subtracted library (MR_K_Y0) was constructed as follows: PCR amplified cDNA inserts from previous library clones from which 3' ESTs had been derived were used as a driver for a hybridization with the normalized whole eye library in the form of single stranded circles. The remaining single stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the MR_K_Y0 library. This procedure has been previously described (Bonaldo, Lennon, and Soares, Genome Research 6: 791-806 1996)."
BASE COUNT 91 a 113 c 105 g 99 t
ORIGIN

Query Match 5.1%; Score 200.6; DB 10; Length 409;
Best Local Similarity 85.5%; Pred. No. 2.2e-58;
Matches 289; Conservative 0; Mismatches 49; Indels 41; Gaps 0;

Qy 1 atgactctctgaaaaaaccctatgaatactatgaaaatgttctatgaactcctcc 60
Db 348 ATGCCTTCTCTGAAAAAATCCTCGTGCCTCTGAGTAGGAAGATCTTTATACTCCTCCTCC 289

Qy 61 ctcccgcaggcccatgctgaaaatcttcctcgaggcgattcctcccgattttgcgatt 120
Db 288 CTCCTGCAGGCCATCATCATGAGAACCTCCTCCTGCATCCTTCCTTCCTGAGTTCCCATCTC 229

Qy 121 tctcctcaaaaaaattttctgtatgaacaacaaaaaaaacgattcctcactatgctat 180
Db 228 TCTCCTCAGGGAAGCTTTCTACGAAGATACCACAGCGGAACACTTCCTCACATACCGCTAT 169

Qy 181 gatgacgaacctcaaaaaaacactgttcaatgaaaacaaaaatgaaacgaatgaatct 240
Db 168 GATGATCAGACCTCAAGAAATGCTCACTCCGATGAACACAAAGATGCCAACTGGCATGCCT 109

Qy 241 tcaaaqcactgaaatgaatgctccccgacctatgcgaaaaaaaaatgatattctctgaa 300
Db 108 TGGGAGACTGGAGTGACTGCTCTCGGACCTGTGGGCGAGCACCTCTATTCTCTGCGA 49

Qy 301 atgatttgactgaaaaaattatgaaaaaaaaaat 338
Db 48 AGATCTCTGACTCGAACGAACTC AAGCACACAACAT 11

RESULT 15
BI526618
LOCUS BI526618 581 bp mRNA linear EST 29-AUG-2001
DEFINITION 604805125F1 NIH_MGC_122 Homo sapiens cDNA clone IMAGE:5201199 5',

```
            mRNA sequence.
ACCESSION   BI524618
VERSION     BI524618.1  GI:15348410
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 581)
  AUTHORS   NIH_MGC_122;.[illegible].
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11504  row: b  column: 16
            High quality sequence stop: 577.
FEATURES             Location/Qualifiers
     source          1..581
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5201199"
                     /clone_lib="NIH_MGC_122"
                     /lab_host="DH10B"
                     /note="Organ: pooled lung and spleen; Vector: pCMV-SPORT6;
                     Site_1: NotI; Site_2: EcoRV (destroyed); RNA source
                     anonymous post-24 week female lung, 24 week female
                     spleen, and 20-22 week male spleen.  Library is oligo-dT
                     primed and directionally cloned (EcoRV site is destroyed
                     upon  cloning).  Average insert size 1.4 kb, insert size
                     range 1-3 kb. Library is normalized and enriched for
                     full-length clones and was  constructed by C. Gruber
                     (Invitrogen).  Research Genetics tracking  code 026. Note:
                     this is a NIH_MGC Library."
BASE COUNT      140 a    151 c    159 g    131 t
ORIGIN


  Query Match             5.0%;  Score 256;  DB 16;  Length 581;
  Best Local Similarity   72.6%;  Pred. No. 6.9e-57;
  Matches  331;  Conservative    0;  Mismatches  125;  Indels    0;  Gaps    0;

Qy     186 ccaqacctcaaqaaaqcactattcaqatqaaaacaaaqatqaqcaactqqqatqcttqqqq 245
           || ||  || || | | | |  || || || |||   || ||   |||||||| ||||
Db     125 CCTGAGTTGCAGGACCGCACGCTCCGAGGAGGACCGGGACGGCCTATGGGATGCCTGGGG 184

Qy     246 cgactqqaqtqactqctcccqqacctqtqqqqqaqqaqcatcatattctctqcqqaqatq 305
           |   |||||||| |||||  || |||||  |||||||| || || || |||||| || | ||
Db     185 CCCATGGAGTGAATGCTCACGCACCTGCGGGGGAGGGGCCTCCTACTCTCTGAGGCGCTG 244

Qy     306 tttqactqqaaqqaatttqtqaaqqqcaqaacattcqqtacaaqacatqcaqccaatcatqa 365
             |||   | | ||  ||||||||    || || || ||||  ||||||||  |||   ||
Db     245 CCTGAGCAGCAAGAGCTGTGAAGGAAGAAATATCCGATACAGAACATGCAGTAATATGTGGA 304

Qy     366 ctqccctccaqatqcaaaaqatttcaqaqccccaqcaqtuctcaqcctacaatqatatcca 425
           |||||| |||| |||  |||||| |||| ||||| ||||||  | |||||||||| |
Db     305 CTGCCCACCACAAGCAGGTGATTCCGAGCTCAGCAATGCTCAGCTCATAATGATGTCAA 364

Qy     426 qtatcaqqqqcattactatqaatqqcttccacqatataatqatcctqctqccccqtqtqc 485
           | | || || || |  |||||||||||||||    | || ||| |||    ||| ||| |
Db     365 GCACCATGCCCAGTTTATGAATGGCTTCCTGTGTCTAATGACCCTGACAACCCATGTTC 424

Qy     486 actcaaqtqtcatqcacaaqtacaaaactqqtqqtqqaqctqqcacctaaqqtactqqa 545
           |||||||| || ||  |||| || | ||| || |  || |.|||||||||  | ||
Db     425 ACTCAAGTGCCAAGCACAAGTGAACAAGCCCGGTGTGGAACTAGCACCTAAGGTCTTAGA 484

Qy     546 tqqaactqqttqcaqtqcqqactcctqqacatatqtatcaatqqcatctqtqaqqcaat 605
```

```
           ||| || |||||| | || || || |||| |||| |||||||| | || ||    ||
Db     485 TGGTAGGCGTTGCTATACAGAATGTTTGGATATGTGCATCAGTGGTTTTATGCCAAATTGT 544

Qy     606 qqqctqcqatcqqcaactqqqaaqcaatqccaaqqa 641
           |||||||||| || |||||||||| | ||||||
Db     545 TGGCTGCGATCACCAGCTGGGAAGCACCGTCAAGGA 580


Search completed: August 19, 2002, 04:01:11
Job time: 4186 sec
```